Two-photon cooperative decay in a cavity in the presence of a thermalized electromagnetic field

N. A. Enaki*) and M. A. Makoveĭ

Institute of Applied Physics, Moldavian Academy of Sciences, MD-2028 Kishinev, Republic of Moldova
(Submitted 15 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 1153–1167 (April 1999)

We study the cooperative two-photon spontaneous decay of an excited atomic system in a microcavity whose size is of the order of several wavelengths of atomic radiation. We show that a thermalized electromagnetic field in the microcavity strongly affects the two-photon cooperative spontaneous emission of radiation. The increase in the rate of spontaneous cooperative decay is due to the presence of a small number of thermalized photons in a microcavity mode. At low temperatures, the two-photon absorption probability is found to be a linear function of the two-photon flux, and photon superbunching is observed. © *1999 American Institute of Physics.* [S1063-7761(99)00104-3]

1. INTRODUCTION

As is known, an excited atom decays spontaneously when an emitter interacts with the vacuum modes of the electromagnetic field. The spontaneous decay of a single atom in a microcavity differs substantially from decay in free space. Since the microcavity is small (its dimensions exceed the wavelength of the radiation only severalfold), spontaneous decay is facilitated substantially when there is resonance between the atomic transition and a microcavity mode and is hindered appreciably when the atomic transition frequency differs from the resonant frequency.[1,2] Such behavior of an atom in a microcavity can be explained by the fact that near the transition frequency the rate of spontaneous decay is proportional to the electromagnetic field mode density.

The study of two-photon light generation has attracted much attention. The various aspects of experimental observation and theoretical treatment of spontaneous decay with respect to the dipole-forbidden transition $|2S\rangle \rightarrow |1S\rangle$ in hydrogen-like and helium-like atoms are discussed, e.g., in Ref. 3. The first report of an experimental observation of two-photon coherent light generation in which excited Li atoms were used was made by Nikolaus *et al.*[4] Recent experiments[5] involving Rydberg atoms have demonstrated the real possibility of building a two-photon micromaser. Also of interest are the observations of spontaneous and thermally stimulated two-photon transitions in the microwave range of the emission spectrum of the Rb atom in resonance fluorescence.[6] A theoretical description of cooperative generation of a light pulse in the two-photon spontaneous decay of atoms via dipole-forbidden transitions can be found in Ref. 7, where it is shown that the two-photon superradiance pulse demonstrates interesting quantum behavior in the absorption process. In Refs. 8, attention is also drawn to the properties of superradiance of the two photons generated in the spontaneous decay of two atoms. Two-photon emission and absorption in the presence of a thermalized electromagnetic field in free space is studied in Ref. 9, where an expression for the time dependence of the half-difference of level populations is derived.

It is well known that the rate of spontaneous decay of Rydberg atoms in microcavities is much higher than in free space.[10–13] Since in a microcavity the rate of two-photon decay in a three-level cascade system increases substantially, it is of interest to study in such a setting the cooperative emission of radiation by Rydberg atoms in microcavities. In the present paper we study the cooperative decay of an ensemble of such atoms with a cascade pattern of the levels, where at a finite temperature the intermediate level is arbitrarily offset from resonance with a microcavity mode. In this case the intermediate level is essentially vacant, and the fact that it lies between the excited and ground states leads to a sizable increase in the two-photon cooperative transition amplitude.[5] We derive exact equations for two atoms separated by a distance r_{21} in the microcavity and participating in two-photon cooperative decay. We also study the temporal behavior of these emitters. The cooperative behavior of the concentrated ensemble of emitters in a two-photon decay process is investigated with the quantum fluctuations of the number of excited atoms ignored. We show that a small number of thermalized photons in the microcavity modes increases the cooperative spontaneous decay rate. Since the two-photon absorption probability is proportional to the second-order correlation function, $w \propto \langle a^{\dagger} a^{\dagger} a a \rangle$, the function is proportional to the two-photon flux (or the number of generated pairs of photons) Moreover, in the cooperative two-photon spontaneous decay process, the second-order correlation function remains greater than the square of the first-order correlation function, which means that pairs of photons are highly correlated. In other words, the generated photons experience superbunching. As the temperature grows, superbunching disappears, and only bunching of the flux of thermalized photons remains.

The plan of the paper is as follows. In Sec. 2 we derive the master equation for an arbitrary operator of the atomic subsystem. For one- and two-atom systems, we obtain exact

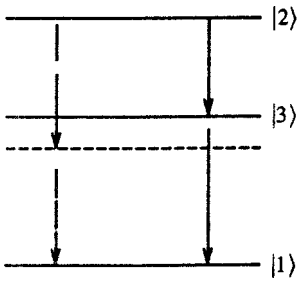

FIG. 1. Diagram of one-photon and two-photon transitions. Solid and broken vertical lines designate one-photon cascade and dipole-forbidden transitions, respectively.

equations that allow for quantum fluctuations of the inversion operator. In Sec. 3 we present the solutions for these differential equations, which make it possible to assess electromagnetic field fluctuations.

2. INTERACTION OF A THREE-LEVEL EXCITED ATOMIC SYSTEM AND A THERMALIZED ELECTROMAGNETIC FIELD IN A MICROCAVITY

We examine an inverted three-level cascade configuration interacting with the modes of an electromagnetic field in a microcavity (see Fig. 1). Since the microcavity is slightly open, the Hamiltonian of such a system has the form

$$H=\sum_k \hbar\omega_k a_k^\dagger a_k+\int_{-\infty}^{\infty} d\omega\,\hbar\omega b_\omega^\dagger b_\omega+\sum_{\alpha=1}^{3}\sum_{j=1}^{N}\hbar\omega_\alpha U_{j\alpha}^{\alpha}$$
$$+i\hbar\sum_k\int_{-\infty}^{\infty} d\omega\,\kappa(\omega)[b_\omega a_k^\dagger-a_k b_\omega^\dagger]$$
$$+i\sum_k\sum_{j=1}^{N}\sum_{\beta=1}^{2}(\mathbf{d}_{3\beta}\cdot\mathbf{g}_k)(a_k^\dagger\exp\{-i\,\mathbf{k}\cdot\mathbf{r}_j\}$$
$$-\text{H.c.})\,(U_{j3}^{\beta}+U_{j\beta}^{3}). \qquad (1)$$

Here $\hbar\omega_\alpha$ ($\alpha=1,2,3$) is the energy of level α, $\mathbf{d}_{3\beta}$ is the dipole moment of the transition between levels $|3\rangle$ and $|\beta\rangle$ ($\beta=1,2$), $a_k^\dagger$ (a_k) is the Bose creation (annihilation) operator for the electromagnetic field inside the microcavity, $b_\omega^\dagger$ (b_ω) is the creation (annihilation) operator outside the cavity, $\mathbf{g}_k=\sqrt{2\pi\hbar\omega_k/V}\,\mathbf{e}_\lambda$, with $\mathbf{e}_\lambda$ the polarization vector of a photon ($\lambda=1,2$) of frequency ω_k and V the cavity volume, $\kappa(\omega)$ is the coupling constant between the external modes and the internal modes of the microcavity, N is the number of atoms in the microcavity, and $U_{j\beta}^{3}=c_{j3}^\dagger c_{j\beta}$ is the corresponding transition operator between levels $|3\rangle$ and $|\beta\rangle$ of the jth atom. Here $c_{j3}^\dagger$ and $c_{j\beta}$ are Fermi operators.

In Eq. (1), the first and second terms describe the electromagnetic field inside the cavity, and the third term is the Hamiltonian of a free atomic system. The fourth term accounts for the interaction of the field inside the cavity with the external field and depends on the reflection coefficient of the cavity walls. Since the atoms are inside the cavity, the fourth term corresponds to the interaction of emitters and microcavity modes.

The operators of the atomic subsystem and electromagnetic field obey the commutation relations

$$[U_{j\beta}^{\alpha},U_{l\alpha'}^{\beta'}]=\delta_{jl}[\delta_{\beta\beta'}U_{j\alpha'}^{\alpha}-\delta_{\alpha\alpha'}U_{j\beta}^{\beta'}],$$
$$[a_k,a_{k'}^\dagger]=\delta_{kk'},\quad [b_\omega,b_{\omega'}^\dagger]=\delta(\omega-\omega').$$

Below we study the temporal behavior of an operator of the atomic subsystem, $O(t)$, in the process of spontaneous decay inside the microcavity. Using the Hamiltonian (1), we derive the Heisenberg equation for the average value of the operator $O(t)$:

$$\frac{d\langle O(t)\rangle}{dt}=\frac{i}{\hbar}\sum_{\alpha=1}^{3}\sum_{j=1}^{N}\hbar\omega_\alpha\langle[U_{j\alpha}^{\alpha}(t),O(t)]\rangle$$
$$-\sum_k\sum_{j=1}^{N}\sum_{\beta=1}^{2}\frac{\mathbf{g}_k\cdot\mathbf{d}_{3\beta}}{\hbar}[\langle a_k^\dagger[U_{j\beta}^{3}(t)$$
$$+U_{j3}^{\beta}(t),O(t)]\rangle\exp\{-i\mathbf{k}\cdot\mathbf{r}_j\}+\text{H.c.}]. \qquad (2)$$

The formal solutions for the operators a_k and $a_k^\dagger$ inside the cavity can be obtained from the Heisenberg equations for operators inside and outside the microcavity (see the Appendix):

$$a_k(t)=A_k^f(t)+a_k^s(t),\quad a_k^\dagger(t)=[a_k(t)]^\dagger, \qquad (3)$$

where

$$A_k^f(t)=a_k(0)\exp\{-i[\omega_k-i\Gamma(\omega_k)]t\}$$
$$+\int_{-\infty}^{\infty}d\omega\,\kappa(\omega)b_\omega(0)$$
$$\times\frac{\exp\{-i\omega t\}-\exp\{-i[\omega_k-i\Gamma(\omega_k)]t\}}{\Gamma(\omega_k)+i(\omega_k-\omega)}, \qquad (4)$$

$$a_k^s(t)=\sum_{j=1}^{N}\sum_{\beta=1}^{2}\frac{\mathbf{d}_{3\beta}\cdot\mathbf{g}_k}{\hbar}\exp\{-i\mathbf{k}\cdot\mathbf{r}_j\}$$
$$\times\int_0^t d\tau\exp\{-i[\omega_k-i\Gamma(\omega_k)]\tau\}$$
$$\times[U_{j\beta}^{3}(t-\tau)+U_{j3}^{\beta}(t-\tau)].$$

Clearly, after (3) and (4) have been substituted into Eq. (2), the right-hand side of this equation acquires a dependence (via A_k^f) on the free operators of the electromagnetic field inside and outside the cavity. We examine the decay of an inverted subsystem of Rydberg atoms at finite temperature, so that the modes inside the cavity and those outside it are partially occupied by thermalized photons. Hence, in studying the dynamics of an inverted atomic subsystem, we can eliminate the free operators of the electromagnetic field inside and outside the cavity by the Bogolyubov method:[14]

$$\frac{d\langle O(t)\rangle}{dt}=i\sum_{\alpha=1}^{3}\sum_{j=1}^{N}\omega_\alpha\langle[U_{j\alpha}^{\alpha}(t),O(t)]\rangle$$
$$-\sum_k\sum_{j,l=1}^{N}\sum_{\beta,\gamma=1}^{2}\frac{(\mathbf{g}_k\cdot\mathbf{d}_{3\beta})(\mathbf{g}_k\cdot\mathbf{d}_{3\gamma})}{\hbar^2}$$
$$\times\Big[\exp\{-i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_l)\}\int_0^t d\tau$$

$$\times\exp\{i[\omega_k+i\Gamma(\omega_k)]\tau\}([n(k)+1]\langle(U^3_{l\gamma}(t-\tau)+U^\gamma_{l3}(t-\tau))[U^3_{j\beta}(t)+U^\beta_{j3}(t),O(t)]\rangle-n(k)\times\langle[U^3_{j\beta}(t)+U^\beta_{j3}(t),O(t)](U^3_{l\gamma}(t-\tau)+U^\gamma_{l3}(t-\tau))\rangle)+\text{H.c.}\Big]. \quad (5)$$

Allowing for the Heisenberg equations for the atomic operators, we can write the solution of the equation for the operator $U^3_{l\gamma}(t-\tau)$ with respect to τ in the form

$$U^3_{l\gamma}(t-\tau)=U^3_{l\gamma}(t)\exp\{-i\omega_{3\gamma}\tau\}+\sum_k\sum_{\eta=1}^{2}\frac{\mathbf{g}_k\cdot\mathbf{d}_{3\eta}}{\hbar}\times\int_0^\tau d\theta\exp(-i\omega_{3\gamma}\theta)(a^\dagger_k(t-\tau+\theta)\times\exp\{-i\mathbf{k}\cdot\mathbf{r}_l\}-\text{H.c.})U^\eta_{l\gamma}(t-\tau+\theta). \quad (6)$$

Inserting this solution into Eq. (5), we obtain

$$\frac{d\langle O(t)\rangle}{dt}=i\sum_{j=1}^{N}\sum_{\alpha=1}^{3}\omega_\alpha\langle[U^\alpha_{j\alpha}(t),O(t)]\rangle-\frac{1}{\hbar^2}\sum_k\sum_{j,l=1}^{N}\Bigg[\exp\{-i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_l)\}[1+n(k)]\times\Bigg((\mathbf{g}_k\cdot\mathbf{d}_{31})^2\frac{\Gamma(\omega_k)+i[\omega_k-\omega_{31}]}{\Gamma^2(\omega_k)+[\omega_k-\omega_{31}]^2}\times\langle U^3_{l1}(t)[U^1_{j3}(t),O(t)]\rangle+(\mathbf{g}_k\cdot\mathbf{d}_{32})^2\frac{\Gamma(\omega_k)+i(\omega_k-\omega_{23})}{\Gamma^2(\omega_k)+(\omega_k-\omega_{23})^2}\langle U^2_{l3}(t)\times[U^3_{j2}(t),O(t)]\rangle\Bigg)-n(k)\exp\{-i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_l)\}\Bigg((\mathbf{g}_k\cdot\mathbf{d}_{31})^2\frac{\Gamma(\omega_k)+i(\omega_k-\omega_{31})}{\Gamma^2(\omega_k)+(\omega_k-\omega_{31})^2}\times\langle[U^1_{j3}(t),O(t)]U^3_{l1}(t)\rangle+(\mathbf{g}_k\cdot\mathbf{d}_{32})^2\frac{\Gamma(\omega_k)+i(\omega_k-\omega_{23})}{\Gamma^2(\omega_k)+(\omega_k-\omega_{23})^2}\times\langle[U^3_{j2}(t),O(t)]U^2_{l3}(t)\rangle\Bigg)+\text{H.c.}\Bigg]+I^{(4)}, \quad (7)$$

where

$$I^{(4)}=-\sum_{k_1,k_2}\sum_{j,l=1}^{N}\sum_{\gamma,\beta,\eta=1}^{2}\frac{(\mathbf{g}_{k_1}\cdot\mathbf{d}_{3\beta})(\mathbf{g}_{k_1}\cdot\mathbf{d}_{3\gamma})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{3\eta})}{\hbar^3}\times\Bigg[\exp\{-i\mathbf{k}_1\cdot(\mathbf{r}_j-\mathbf{r}_l)\}\int_0^t d\tau\exp\{i[\omega_{k_1}+i\Gamma(\omega_{k_1})]\tau\}\times\Bigg([1+n(k_1)]\Bigg\langle\Bigg\{\int_0^\tau d\theta\exp\{-i\omega_{3\gamma}\theta\}[a^\dagger_{k_2}(t-\tau+\theta)\exp\{-i\mathbf{k}_2\cdot\mathbf{r}_1\}-\text{H.c.}]U^\eta_{l\gamma}(t-\tau+\theta)+\text{H.c.}\Bigg\}\times[U^3_{j\beta}(t)+U^\beta_{j3}(t),O(t)]\Bigg\rangle-n(k_1)\Bigg\langle[U^3_{j\beta}(t)+U^\beta_{j3}(t),O(t)]\Bigg\{\int_0^\tau d\theta\exp\{-i\omega_{3\gamma}\theta\}\times[a^\dagger_{k_2}(t-\tau+\theta)\exp\{-i\mathbf{k}_2\cdot\mathbf{r}_1\}-\text{H.c.}]\times U^\eta_{l\gamma}(t-\tau+\theta)+\text{H.c.}\Bigg\}\Bigg\rangle\Bigg)+\text{H.c.}\Bigg]. \quad (8)$$

The second term on the right-hand side of (7) accounts for one-photon transitions of type $|2\rangle\to|3\rangle$ or $|3\rangle\to|1\rangle$, and the third term accounts for higher-order transitions.

Since we are only interested in two-photon transitions, we ignore one-photon emission by proper selection of the offset of the intermediate level from resonance. More precisely, the offset $\Delta=|\omega_k-\omega_{23}|$ ($\Delta=|\omega_k-\omega_{31}|$) must be much greater than the damping factor $\Gamma(k)$ of the microcavity. Hence the only mechanism by which the system can go into the ground state is two-photon decay. Clearly, in this case the level $|3\rangle$ is almost vacant, and we can eliminate it from Eq. (8). As an example, we eliminate it from the first term on the right-hand side of Eq. (8). The result is

$$I^{(4)}_1=\sum_{k_1,k_2}\sum_{j,l=1}^{N}\sum_{\gamma,\beta,\eta=1}^{2}\frac{(\mathbf{g}_{k_1}\cdot\mathbf{d}_{3\beta})(\mathbf{g}_{k_1}\cdot\mathbf{d}_{3\gamma})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{3\eta})}{\hbar^3}\times\exp\{-i\mathbf{k}_1\cdot(\mathbf{r}_j-\mathbf{r}_l)\}\int_0^t d\tau\exp\{i[\omega_{k_1}+i\Gamma(\omega_{k_1})]\tau\}[1+n(k_1)]\int_0^\tau d\theta\exp\{-i\omega_{3\gamma}\theta\}\langle a_{k_2}(t-\tau+\theta)\exp\{i\mathbf{k}_2\cdot\mathbf{r}_1\}U^\eta_{l\gamma}(t-\tau+\theta)[U^3_{j\beta}(t),O(t)]\rangle. \quad (9)$$

Since $U^3_{j\beta}(t)=c^\dagger_{3j}(t)c_{j\beta}(t)$, using the solution of the Heisenberg equation for the Fermi operator $c^\dagger_{3j}(t)$,

$$c^\dagger_{3j}(t)=c^\dagger_{3j}(0)\exp\{i\omega_3 t\}+\sum_{k_3}\sum_{\xi=1}^{2}\frac{\mathbf{d}_{3\xi}\cdot\mathbf{g}_{k_3}}{\hbar}\int_0^t d\tau_1\exp\{i\omega_3\tau_1\}\times[a_{k_3}(t-\tau_1)\exp\{i\mathbf{k}\cdot\mathbf{r}_j\}-\text{H.c.}]c^\dagger_{j\xi}(t-\tau_1),$$

we can eliminate it from (9). Inserting this solution into (9) and bearing in mind that $\langle 3|c^\dagger_{j3}(0)=0$ and $c_{j3}(0)|3\rangle=0$, we immediately obtain

$$I_1^{(4)}=\sum_{k_1,k_2,k_3}\sum_{j,l,=1}^{N}\sum_{\beta,\gamma,\eta,\xi=1}^{2}\times\frac{(\mathbf{g}_{k_1}\cdot\mathbf{d}_{3\beta})(\mathbf{g}_{k_1}\cdot\mathbf{d}_{3\gamma})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{3\eta})(\mathbf{g}_{k_3}\cdot\mathbf{d}_{3\xi})}{\hbar^4}\times\exp\{-i\mathbf{k}_1\cdot(\mathbf{r}_j-\mathbf{r}_l)\}\int_0^t d\tau\exp\{i[\omega_{k_1}+i\Gamma(\omega_{k_1})]\tau\}[1+n(k_1)]\int_0^\tau d\theta\exp\{-i\omega_{3\gamma}\theta\}\int_0^t d\tau_1\times\exp\{i\omega_{3\beta}\tau_1\}\exp\{i\mathbf{k}_2\cdot\mathbf{r}_l-i\mathbf{k}_3\cdot\mathbf{r}_j\}\langle a_{k_2}(t-\tau+\theta)a_{k_3}^{\dagger}(t-\tau_1)U_{l\gamma}^{\eta}(t-\tau+\theta)[U_{j\beta}^{\xi}(t-\tau_1),O(t)]\rangle. \tag{10}$$

We now ignore terms of order higher than the fourth with respect to the parameters $g_k d_{31}/\hbar$ and $g_k d_{32}/\hbar$ and write the correlation function as follows:

$$\langle a_{k_2}(t-\tau+\theta)a_{k_3}^{\dagger}(t-\tau_1)\rangle\approx\delta_{k_2k_3}[1+n(k_2)]\times\exp\{i\omega_{k_2}(\tau-\theta-\tau_1)\}\times\exp\{-\Gamma(\omega_{k_2})|\tau-\theta-\tau_1|\}.$$

Next we use an integral representation of the exponentials:

$$\exp\{i\omega_{k_2}(\tau-\theta-\tau_1)\}\exp\{-\Gamma(\omega_{k_2})|\tau-\theta-\tau_1|\}=\frac{1}{\pi}\int_{-\infty}^{\infty}d\omega'\frac{\Gamma(\omega_{k_2})\exp\{i\omega'(\tau-\theta-\tau_1)\}}{(\omega'-\omega_{k_2})^2+\Gamma^2(\omega_{k_2})},$$

$$\exp\{i[\omega_{k_1}+i\Gamma(k_1)]\tau\}=\frac{1}{\pi}\int_{-\infty}^{\infty}d\omega''\frac{\Gamma(\omega_{k_1})\exp\{i\omega''\tau\}}{(\omega''-\omega_{k_1})^2+\Gamma^2(\omega_{k_1})}.$$

Then Eq. (10) becomes

$$I_1^{(4)}\approx\sum_{k_1,k_2}\sum_{j,l=1}^{N}\sum_{\beta,\gamma,\eta,\xi=1}^{2}\times\frac{(\mathbf{g}_{k_1}\cdot\mathbf{d}_{3\beta})(\mathbf{g}_{k_1}\cdot\mathbf{d}_{3\gamma})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{3\eta})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{3\xi})}{\pi^2\hbar^4}\times\exp\{-i(\mathbf{k}_1+\mathbf{k}_2)\cdot(\mathbf{r}_j-\mathbf{r}_l)\}[1+n(k_1)]\times[1+n(k_2)]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega'\,d\omega''\times\frac{\Gamma(\omega_{k_2})}{\Gamma^2(\omega_{k_2})+(\omega'-\omega_{k_2})^2}\frac{\Gamma(\omega_{k_1})}{\Gamma^2(\omega_{k_1})+(\omega''-\omega_{k_1})^2}\times\langle U_{l\gamma}^{\eta}(t)[U_{j\beta}^{\xi}(t),O(t)]\rangle I_1'(\omega',\omega''). \tag{11}$$

Here the function $I_1'(\omega',\omega'')$ can be expressed in terms of temporal integrals, and for long time intervals, $t\to\infty$, we can write

$$I_1'(\omega',\omega'')=\int_0^t d\tau\exp\{i(\omega'+\omega''-\omega_{\eta\gamma})\tau\}\int_0^\tau d\theta\times\exp\{-i(\omega'+\omega_{3\eta})\theta\}\int_0^t d\tau_1\times\exp\{i(\omega_{3\xi}-\omega')\tau_1\}\approx\frac{\pi\delta(\omega'+\omega''-\omega_{\eta\gamma})}{(\omega_{3\eta}+\omega')(\omega_{3\xi}-\omega')}.$$

After integration over time, the integrals with respect to ω' and ω'' in Eq. (11) can easily be evaluated:

$$I_1''=\frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}d\omega'\,d\omega''\,I_1'(\omega',\omega'')\times\frac{\Gamma(\omega_{k_1})}{\Gamma^2(\omega_{k_1})+(\omega''-\omega_{k_1})^2}\frac{\Gamma(\omega_{k_2})}{\Gamma^2(\omega_{k_2})+(\omega'-\omega_{k_2})^2}\approx\frac{1}{(\omega_{31}-\omega_{k_1})(\omega_{32}+\omega_{k_1})}\frac{2\Gamma}{(\omega_{k_1}+\omega_{k_2}-\omega_{21})^2+(2\Gamma)^2}.$$

In final form the expression for $I_1^{(4)}$ is

$$I_1^{(4)}\approx\sum_{k_1,k_2}\sum_{j,l=1}^{N}\frac{(\mathbf{g}_{k_1}\cdot\mathbf{d}_{31})(\mathbf{g}_{k_1}\cdot\mathbf{d}_{32})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{31})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{32})}{\hbar^4(\omega_{31}-\omega_{k_1})(\omega_{32}-\omega_{k_2})}\times\exp\{-i(\mathbf{k}_1+\mathbf{k}_2)\cdot(\mathbf{r}_j-\mathbf{r}_l)\}[1+n(k_1)]\times[1+n(k_2)]\langle U_{l1}^{2}(t)[U_{j2}^{1}(t),O(t)]\rangle\times\frac{2\Gamma}{(\omega_{k_1}+\omega_{k_2}-\omega_{21})^2+(2\Gamma)^2}. \tag{12}$$

In deriving Eq. (12) we used the rotating wave approximation and assumed that $\Gamma(\omega_{k_1})\approx\Gamma(\omega_{k_2})=\Gamma$, $\omega_{31}-\omega_{k_1}\gg\Gamma$, and $\omega_{32}+\omega_{k_1}\gg\Gamma$. Reasoning in the same manner, we can derive an equation describing the behavior of atoms interacting with a thermalized electromagnetic field in a microcavity:

$$\frac{d\langle O(t)\rangle}{dt}=i\omega_{21}\sum_{j=1}^{N}\langle[R_{zj},O(t)]\rangle+\frac{1}{\hbar^4}\sum_{k_1,k_2}\sum_{j,l=1}^{N}\left[\frac{(\mathbf{g}_{k_1}\cdot\mathbf{d}_{31})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{32})}{\omega_{23}-\omega_{k_2}}-\frac{(\mathbf{g}_{k_1}\cdot\mathbf{d}_{32})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{31})}{\omega_{31}-\omega_{k_2}}\right]^2\Bigg[[1+n(k_1)+n(k_2)]\times\Bigg(\exp\{-i(\mathbf{k}_1+\mathbf{k}_2)\cdot(\mathbf{r}_j-\mathbf{r}_l)\}\times\frac{\langle R_l^{+}[R_j^{-},O(t)]\rangle}{i(\omega_{k_1}+\omega_{k_2}-\omega_{21}+2i\Gamma)}+\exp\{i(\mathbf{k}_1$$

$$+\mathbf{k}_2)\cdot(\mathbf{r}_j-\mathbf{r}_l)\}\frac{\langle[R_j^+,O(t)]R_l^-\rangle}{i(\omega_{k_1}+\omega_{k_2}-\omega_{21}-2i\Gamma)}\Bigg)$$

$$+n(k_1)n(k_2)\Bigg(\exp\{-i(\mathbf{k}_1+\mathbf{k}_2)\cdot(\mathbf{r}_j-\mathbf{r}_l)\}$$

$$\times\frac{\langle R_l^+[R_j^-,O(t)]\rangle}{i(\omega_{k_1}+\omega_{k_2}-\omega_{21}+2i\Gamma)}+\exp\{i(\mathbf{k}_1+\mathbf{k}_2)\cdot(\mathbf{r}_j$$

$$-\mathbf{r}_l)\}\frac{\langle[R_j^+,O(t)]R_l^-\rangle}{i(\omega_{k_1}+\omega_{k_2}-\omega_{21}-2i\Gamma)}\Bigg)\Bigg], \tag{13}$$

where the operators $R_j^+=U_{j1}^2$, $R_j^-=U_{j2}^1$, and $R_{zj}=(U_{j2}^2-U_{j1}^2)/2$ satisfy the commutation relations for spin operators.[7]

To study the time dependence of the operators of the atomic subsystem, we write the following system of equations:

$$\frac{d\langle R_{zj}\rangle}{dt}=-\frac{2}{\hbar^4}\sum_{k_1,k_2}\sum_{j,l=1}^{N}\gamma_{\rm eff}(\mathbf{k}_1,\mathbf{k}_2)\exp\{-i(\mathbf{k}_1+\mathbf{k}_2)\cdot(\mathbf{r}_j$$

$$-\mathbf{r}_l)\}[1+n(k_1)+n(k_2)]\langle R_l^+(t)R_j^-(t)\rangle$$

$$-\frac{4}{\hbar^2}\sum_{k_1,k_2}\gamma_{\rm eff}(\mathbf{k}_1,\mathbf{k}_2)n(k_1)n(k_2)\langle R_{zj}(t)\rangle, \tag{14}$$

$$\frac{d\langle R_i^+(t)R_m^-(t)\rangle}{dt}=\frac{2}{\hbar^4}\sum_{k_1,k_2}\sum_{l=1}^{N}\gamma_{\rm eff}^0(\mathbf{k}_1,\mathbf{k}_2)\Bigg[[1+n(k_1)$$

$$+n(k_2)]\Bigg(\exp\{i(\mathbf{k}_1+\mathbf{k}_2)\cdot(\mathbf{r}_m-\mathbf{r}_l)\}$$

$$\times\frac{\langle R_i^+(t)R_{zm}(t)R_l^-(t)\rangle}{i(\omega_{k_1}+\omega_{k_2}-\omega_{21}-2i\Gamma)}$$

$$-\exp\{-i(\mathbf{k}_1+\mathbf{k}_2)\cdot(\mathbf{r}_i-\mathbf{r}_l)\}$$

$$\times\frac{\langle R_l^+(t)R_{zi}(t)R_m^-(t)\rangle}{i(\omega_{k_1}+\omega_{k_2}-\omega_{21}+2i\Gamma)}\Bigg)\Bigg],$$

where

$$\gamma_{\rm eff}^0(\mathbf{k}_1,\mathbf{k}_2)=\left[\frac{(\mathbf{g}_{k_1}\cdot\mathbf{d}_{31})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{32})}{\omega_{23}-\omega_{k_2}}-\frac{(\mathbf{g}_{k_1}\cdot\mathbf{d}_{32})(\mathbf{g}_{k_2}\cdot\mathbf{d}_{31})}{\omega_{31}-\omega_{k_2}}\right]^2,$$

$$\gamma_{\rm eff}(\mathbf{k}_1,\mathbf{k}_2)=\gamma_{\rm eff}^0(\mathbf{k}_1,\mathbf{k}_2)\frac{2\Gamma}{(2\Gamma)^2+(\omega_{k_1}+\omega_{k_2}-\omega_{21})^2}.$$

Next we limit ourselves to one microcavity mode. We study the kinetics of one, two, or more atoms and take into account the quantum fluctuations in the case of one or two atoms and ignore these fluctuations if the number of atoms is greater than two.

We begin with one atom in the microcavity. Equations (14) yield an equation for the population of the atomic subsystem:

$$\frac{d\langle R_z(t)\rangle}{dt}=-\frac{1}{\tau_0^{(b)}}\langle R^+R^-\rangle-\frac{1}{\tau^{(b)}}\langle R_z\rangle, \tag{15}$$

where

$$\frac{1}{\tau_0^{(b)}}=\frac{4g^4d_0^4}{\Gamma\Delta^2\hbar^4}(1+2n),\quad \frac{1}{\tau^{(b)}}=\frac{8g^4d_0^4n^2}{\Gamma\Delta^2\hbar^4}(1+2n).$$

With the initial condition $\langle R_z(t=0)\rangle=1/2$ we have

$$\langle R_z(t)\rangle=\frac{1}{2}\left[1-\frac{a+1}{a}\left(1-\exp\left\{-\frac{at}{\tau_0^{(b)}}\right\}\right)\right], \tag{16}$$

where

$$a=1+\frac{\tau_0^{(b)}}{\tau^{(b)}}=1+\frac{2n^2}{1+2n}.$$

In the case of two atoms separated by a distance $r_{21}=|\mathbf{r}_2-\mathbf{r}_1|$, the system of equations (14) leads to a complete set of equations for the variables $Z(t)=\langle R_{z1}(t)\rangle+\langle R_{z2}(t)\rangle$, $Y(t)=\langle R_2^+(t)R_1^-(t)\rangle+\langle R_1^+(t)R_2^-(t)\rangle$, and $X(t)=\langle R_{z1}(t)R_{z2}(t)\rangle$:

$$\frac{d}{d\tau}\begin{bmatrix}Z(\tau)\\Y(\tau)\\X(\tau)\end{bmatrix}=\begin{bmatrix}-a & -\cos\theta & 0\\ \cos\theta & -1 & 4\cos\theta\\ -0.5 & 0.5a\cos\theta & -2a\end{bmatrix}$$

$$\times\begin{bmatrix}Z(\tau)\\Y(\tau)\\X(\tau)\end{bmatrix}+\begin{bmatrix}-1\\0\\0\end{bmatrix}, \tag{17}$$

where $\tau=t/\tau_0^{(b)}$, and $\theta=\mathbf{k}_0\cdot\mathbf{r}_{21}$. The characteristic equation of this new system of equations is

$$\lambda^3+(3a+1)\lambda^2+(3a+2a^2-2a\cos^2\theta+\cos^2\theta)\lambda$$

$$+2a^2-2\cos^2\theta-2a^2\cos^2\theta+2a\cos^2\theta=0.$$

Note that this equation can easily be solved by Cramer's method, but in view of the awkwardness of that method we limit ourselves to the approximation $\alpha\approx1$, which corresponds to a small number of thermalized photons in the microcavity mode. Allowing for the initial conditions $Z(t=0)=1$, $Y(t=0)=0$, and $X(t=0)=0.25$, we obtain in this approximation the following solution of the system of equations (17):

$$Z(\tau)=-\frac{4p^2}{1-p^2}\exp\{-2\tau\}+\frac{1-p}{1+p}\exp\{-(1-p)\tau\}$$

$$+\frac{1+p}{1-p}\exp\{-(1+p)\tau\}-1,$$

$$Y(\tau)=-\frac{4p^2}{1-p^2}\exp\{-2\tau\}-\frac{1-p}{1+p}\exp\{-(1-p)\tau\}$$

$$+\frac{1+p}{1-p}\exp\{-(1+p)\tau\}, \tag{18}$$

$$X(\tau)=\frac{1+p^2}{1-p^2}\exp\{-2\tau\}-\frac{1-p}{2(1+p)}\exp\{-(1-p)\tau\}$$
$$-\frac{1+p}{2(1-p)}\exp\{-(1+p)\tau\}+0.25,$$

where $p=\cos\theta$.

When the number of atoms is large, the process of cooperative two-photon decay becomes appreciably stronger. If we ignore fluctuations in the number of particles when the number of atoms is large ($N\gg 1$), we can easily obtain for the atomic inversion operator

$$\frac{d\langle R_z(t)\rangle}{dt}=-\frac{1}{\tau^{(b)}}\langle R_z\rangle-\frac{1}{\tau_0^{(b)}}(j(j+1)-\langle R_z\rangle^2+\langle R_z\rangle), \tag{19}$$

where $j=N/2$. The solution of this equation is

$$\langle R_z(t)\rangle=\frac{1+q}{2}-\frac{c}{2}\tanh\left[\frac{1}{2\tau_r}(t-t_0)\right],$$

where $\tau_r=\tau_0^{(b)}/c$ is the time of cooperative spontaneous decay of the ensemble of atoms,

$$t_0=\tau_r\ln\frac{N-(1+q-c)}{(1+q+c)-N}$$

is the time lag of the pulse of collective emission of a pair of photons in the microcavity, $q=\tau_0^{(b)}/\tau^{(b)}$, and $c=\sqrt{(1+q)^2+4j(j+1)}$.

The Dicke equation (19) describing two-photon cooperative spontaneous decay implies that a thermalized field affects not only the Einstein coefficient $1/\tau^{(b)}$ corresponding to stimulated decay but also the rate of two-photon spontaneous decay, $1/\tau_0^{(b)}$. Clearly, a thermalized field facilitates the process of cooperative two-photon decay [see the expression for $1/\tau_0^{(b)}$ after Eq. (15)]. This constitutes one of the main differences between two-photon dipole-forbidden emission and one-photon cascade cooperative spontaneous emission. Obviously, two-photon cooperative spontaneous emission prevails over stimulated thermalized transition only if $N(1+2n)>n^2$. These estimates suggest that when $n<1$, the term $1/\tau^{(b)}$, which corresponds to induced decay, is negligible in comparison to the term $1/\tau_0^{(b)}$, which corresponds to spontaneous decay. Indeed, at $n=0.3$ we have

$$\frac{1}{\tau^{(b)}}=\frac{8g^4d_0^4}{\Gamma\Delta^2\hbar^4}\times 0.3^2\ll\frac{1}{\tau_0^{(b)}}=\frac{4g^4d_0^4}{\Gamma\Delta^2\hbar^4}[1+2\times 0.3].$$

To conclude this section, we note that with properly selected (by experiments). Rydberg atoms and microcavity mode, the one-photon cascade process can be neglected in comparison to two-photon spontaneous decay. To quench the one-photon cascade transition, the rate of loss of photons from the microcavity, Γ, must be smaller than the offset from resonance, $\Delta=|\omega_c-\omega_{23}|(\Delta=|\omega_c-\omega_{31}|)$, so that $\Gamma^2\ll\Delta^2$ (here ω_c is the microcavity mode frequency). Bearing in mind the condition for applicability of the Born–Markov approximation in second-order perturbation theory, $\Gamma>N/\tau_0^{(b)}$, we obtain the following lower and upper bounds on the photon loss rate:

$$\frac{4g^4d_0^4N}{\Delta^2\hbar^4}(1+2n)<\Gamma^2\ll\Delta^2. \tag{20}$$

Our model can be experimentally implemented for the $|n'S\rangle\rightarrow|(n'-1)S\rangle$ transition with an intermediate level $|(n'-1)P\rangle$ and with allowance for (20). Since we wish to account for the effect of the heat bath on two-photon spontaneous decay, we examine the experimental model of excitation for Rb atoms with $n'=40$ (Ref. 5). The average number n of thermalized photons in the microcavity mode at $T=4$ K is of order $n\approx 0.79$. In the notation of Ref. 5, the conditions (20) become

$$\frac{4\Omega_{if}^4N(1+2n)}{\Delta^2}<\Gamma^2\ll\Delta^2.$$

Then, with the values of the matrix element Ω_{if} and the offset Δ obtained in Ref. 5, we easily establish that

$$1.6\times 10^7N(1+2\times 0.79)<\Gamma^2\ll 1.5\times 10^{15},$$

i.e., the inequalities in (20) hold. Note that in this case the rate of collective two-photon spontaneous decay is 2.6 times the rate of two-photon decay in the absence of thermalized photons in the cavity mode (i.e., at $T=0$).

3. DEPENDENCE OF ELECTROMAGNETIC FIELD FLUCTUATIONS ON ATOMIC INVERSION DYNAMICS

In the absence of atoms inside the microcavity, we can calculate the fluctuations of the electromagnetic field operators:

$$\delta_0^2=\langle a^{\dagger\,2}a^2\rangle-\langle a^\dagger a\rangle^2=n^2.$$

It would be interesting to find the fluctuations in the number of photons of the electromagnetic field that are generated by the excited atomic system in the process of two-photon emission in the microcavity. To do this we introduce a function that accounts for fluctuations of the electromagnetic field in relation to thermalized fluctuations:

$$\delta_r^2=\delta^2-\delta_0^2, \tag{21}$$

where $\delta^2=\langle a^{\dagger 2}(t)a^2(t)\rangle-\langle a^\dagger(t)a(t)\rangle^2$.

Since experimenters often monitor the dynamics of the population difference of the atomic subsystem in the microcavity,[10] in this section we express the electromagnetic field fluctuations δ_r^2 in terms of the kinetics of atomic population inversion. To simplify this problem, we can eliminate the virtual levels $c_{j3}^\dagger(t)$ and $c_{j3}(t)$ from the Hamiltonian (1), since as noted in Sec. 2, these levels are almost vacant. After this is done, we obtain a formula for the effective Hamiltonian describing the interaction of the atomic subsystem and a single microcavity mode at $T\neq 0$:

$$H^{\text{eff}}=\hbar\omega_{21}R_z+\hbar\omega_c J_z+2\hbar\chi(J^+R^-+R^+J^-)+\int_{-\infty}^{\infty}d\omega\,\hbar\omega b_\omega^\dagger b_\omega+i\hbar\times\int_{-\infty}^{\infty}d\omega\,\kappa(\omega)(b_\omega a^\dagger-ab_\omega^\dagger),\tag{22}$$

where $\chi=2g^2d_{31}d_{32}/\Delta\hbar^2$. The operators $J^+=a^{\dagger 2}/2$, $J^-=a^2/2$, and $J_z=(a^\dagger a+1/2)/2$ belong to the SU(1,1) algebra and satisfy the commutation relations

$$[J^+,J^-]=-2J_z,\quad [J_z,J^\pm]=\pm J^\pm.$$

Let $G^{(1)}(t)=\langle a^\dagger(t)a(t)\rangle$. Then

$$\frac{dG^{(1)}(t)}{dt}=\frac{d\langle a^\dagger(t)a(t)\rangle}{dt}=\left\langle\frac{da^\dagger(t)}{dt}a(t)+a^\dagger(t)\frac{da(t)}{dt}\right\rangle.\tag{23}$$

Inserting the Heisenberg equation of the operators of the cavity's electromagnetic field into Eq. (23), we obtain

$$\frac{dG^{(1)}(t)}{dt}=4i\chi\langle R^+(t)J^-(t)-J^+(t)R^-(t)\rangle-2\Gamma\langle a^\dagger(t)a(t)\rangle+\int_{-\infty}^{\infty}d\omega\,\kappa(\omega)\times\langle b_\omega^\dagger(0)a(t)\rangle\exp\{i\omega t\}+\int_{-\infty}^{\infty}d\omega\,\kappa(\omega)\times\langle a^\dagger(t)b_\omega(0)\exp\{-i\omega t\}.\tag{24}$$

Eliminating the heat-bath operators, we obtain the following equation for the first-order correlation function:

$$\frac{dG^{(1)}(t)}{dt}=-2\Gamma G^{(1)}(t)+4i\chi\langle R^+(t)J^-(t)-J^+(t)R^-(t)\rangle+2\Gamma n.$$

In the Born–Markov approximation $dG^{(1)}(t)/dt\ll\Gamma G^{(1)}(t)$, we can express the function $G^{(1)}(t)$ in terms of the atomic inversion operator:

$$G^{(1)}(t)=n-\frac{1}{\Gamma}\frac{d\langle R_z(t)\rangle}{dt}.\tag{25}$$

Reasoning along the same lines, we find the second-order correlation function $G^{(2)}(t)=\langle a^{\dagger 2}(t)a^2(t)\rangle$:

$$\frac{dG^{(2)}(t)}{dt}=\frac{d\langle a^{\dagger 2}(t)a^2(t)\rangle}{dt}=\left\langle\frac{da^{\dagger 2}(t)}{dt}a^2(t)+a^{\dagger 2}(t)\frac{da^2(t)}{dt}\right\rangle.\tag{26}$$

Taking into account the Heisenberg equation for the operator $a^{\dagger 2}(t)$,

$$\frac{da^{\dagger 2}}{dt}=4i\omega_cJ^+-4\Gamma J^++8i\chi R^+J_z+\int_{-\infty}^{\infty}d\omega\,\kappa(\omega)b_\omega^\dagger(0)a^\dagger(t)\exp\{i\omega t\},\tag{27}$$

and inserting it into Eq. (26) in the Born–Markov approximation $dG^{(2)}(t)/dt\ll\Gamma G^{(2)}(t)$, we find that $G^{(2)}(t)$ and $G^{(1)}(t)$ are linked through the relationship

$$G^{(2)}(t)=\frac{4i\chi}{\Gamma}[\langle R^+J_zJ^-\rangle-\langle R^-J^+J_z\rangle]+2nG^{(1)}(t).\tag{28}$$

Now, weakening the correlation functions

$$\langle R^+J_zJ^-\rangle\approx\langle R^+J^-\rangle\langle J_z\rangle,\quad \langle R^-J^+J_z\rangle\approx\langle R^-J^+\rangle\langle J_z\rangle$$

for a large number of excited atoms and noting that

$$\langle J_z(t)\rangle=\frac{1}{2}\left[G^{(1)}(t)+\frac{1}{2}\right]=\frac{1}{2}\left[n+\frac{1}{2}-\frac{1}{\Gamma}\frac{d\langle R_z(t)\rangle}{dt}\right],$$

we can express the second-order correlation function in terms of the atomic subsystem inversion operator:

$$G^{(2)}(t)=2n^2-\left[3n+\frac{1}{2}\right]\frac{1}{\Gamma}\frac{d\langle R_z(t)\rangle}{dt}+\left[\frac{1}{\Gamma}\frac{d\langle R_z(t)\rangle}{dt}\right]^2.\tag{29}$$

Note that since the two-photon absorption probability $w\propto\langle a^{\dagger 2}a^2\rangle=G^{(2)}(t)$, at low-temperatures it is proportional to the two-photon flux Φ_0:

$$w\propto\frac{d\langle R_z(t)\rangle}{dt}\propto\Phi.$$

We note that the probability w also depends on the square of the two-photon flux, but this dependence is ignored in our approximation. Note that for one-photon superradiance, the function $G^{(2)}$ is proportional to the square of the one-photon flux, or Φ^2. This occurs because $\alpha^\dagger\sim R^+$ for one-photon emission, while for two-photon emission we have $a^{\dagger 2}\sim R^+$. Hence

$$w\propto\langle R^{+\,2}R^{-\,2}\rangle\propto\left\langle\left[\frac{dR_z}{dt}\right]^2\right\rangle$$

for one-photon superradiance, and

$$w\propto\langle R^+R^-\rangle\propto\frac{d\langle R_z\rangle}{dt}$$

for two-photon superradiance.

Now we can easily derive a formula for the relative fluctuations of the electromagnetic field inside the microcavity:

$$\delta_r^2=-\left(n+\frac{1}{2}\right)\frac{1}{\Gamma}\frac{d\langle R_z(t)\rangle}{dt}.\tag{30}$$

This implies that in each decay event, photons are generated in pairs and the emission intensity becomes proportional to N^2, while the second-order correlation function for the photons remains much greater than the square of the

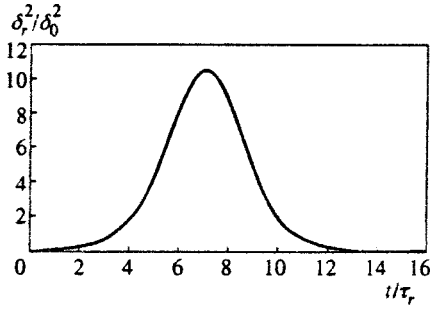

FIG. 2. Dependence of δ_r^2/δ_0^2 on t/τ_r at $N=1500$, $\Delta/2\pi=39$ MHz, $\Gamma=2\times10^6\ \mathrm{s}^{-1}$, $\omega_c/2\pi=68.4$ GHz, and $\Omega=7\times10^5\ \mathrm{s}^{-1}$.

first-order correlation function. In this case, at low temperatures and for large numbers of atoms, we can speak of photon superbunching, i.e. $\delta_r^2/\delta_0^2 \gg 1$ (Fig. 2).

4. CONCLUSION

The main object of this paper was to study the cooperative two-photon spontaneous decay of excited Rydberg atoms in a microcavity in the presence of a thermalized field. We have found that a thermalized electromagnetic field boosts two-photon spontaneous emission and that this emission prevails over stimulated emission. Moreover, we have studied electromagnetic field fluctuations generated by the excited atomic subsystem in relation to fluctuations of the thermalized field in the microcavity. Finally, we have established, in the Born–Markov approximation, that these fluctuations are linear functions of the generated two-photon flux.

APPENDIX

We show how to eliminate the operators of the thermalized electromagnetic field and the electromagnetic field of the microcavity so as to obtain Eq. (5). We start by writing the Heisenberg equations for the operators of the electromagnetic field,

$$\frac{da_k(t)}{dt}=-i\omega_k a_k(t)+\int_{-\infty}^{\infty}d\omega\,\kappa(\omega)b_\omega(t)+\sum_{j=1}^{N}\sum_{\beta=1}^{2}\frac{\mathbf{g}_k\cdot\mathbf{d}_{3\beta}}{\hbar}\exp\{-i\mathbf{k}\cdot\mathbf{r}_j\}\times[U_{j\beta}^3(t)+U_{j3}^{\beta}(t)], \tag{A1}$$

and of the thermalized electromagnetic field,

$$b_\omega(t)=b_\omega(0)\exp\{-i\omega t\}-\sum_k\int_0^t d\tau\exp\{-i\omega\tau\}a_k(t-\tau)\kappa(\omega). \tag{A2}$$

Substituting (A2) into (A1), we obtain the Markov approximations of the equations for the operators of the electromagnetic field of the microcavity:

$$\frac{da_k(t)}{dt}=-\sum_{k'}\left(i\omega_{k'}\delta_{kk'}+\int_0^\infty d\omega\,\kappa^2(\omega)\frac{\varepsilon+i(\omega_{k'}-\omega)}{\varepsilon^2+(\omega_{k'}-\omega)^2}\right)a_{k'}(t)+\int_{-\infty}^{\infty}d\omega\,\kappa(\omega)b_\omega(0)\exp\{-i\omega t\}+\sum_{j=1}^{N}\sum_{\beta=1}^{2}\frac{\mathbf{g}_k\cdot\mathbf{d}_{3\beta}}{\hbar}\exp\{-i\mathbf{k}\cdot\mathbf{r}_j\}[U_{j\beta}^3(t)+U_{j3}^{\beta}(t)]. \tag{A3}$$

If we ignore the contribution of adjacent modes to the redistribution of frequencies and to microcavity losses, we obtain

$$\tilde{\omega}_k=\omega_k-P\int_{-\infty}^{\infty}d\omega\frac{|\kappa(\omega)|^2}{\omega-\omega_k},\quad \Gamma(\omega_k)=\pi|\kappa(\omega_k)|^2.$$

In this approximation, the formal solution for the operators of the microcavity's electromagnetic field is

$$a_k(t)=A_k^f(t)+a_k^s(t),\quad a_k^\dagger(t)=[a_k(t)]^\dagger. \tag{A4}$$

Here

$$A_k^f(t)=a_k(0)\exp\{-i[\omega_k-i\Gamma(\omega_k)]t\}+\int_{-\infty}^{\infty}d\omega\,\kappa(\omega)b_\omega(0)\times\frac{\exp\{-i\omega t\}-\exp\{-i[\omega_k-i\Gamma(\omega_k)]t\}}{\Gamma(\omega_k)+i(\omega_k-\omega)},$$

$$a_k^s(t)=\sum_{j=1}^{N}\sum_{\beta=1}^{2}\frac{\mathbf{d}_{3\beta}\cdot\mathbf{g}_k}{\hbar}\exp\{-i\mathbf{k}\cdot\mathbf{r}_j\}\int_0^t d\tau\exp\{-i[\omega_k-i\Gamma(\omega_k)]\tau\}[U_{j\beta}^3(t-\tau)+U_{j3}^{\beta}(t-\tau)],$$

where we have introduced the notation $\tilde{\omega}_k=\omega_k$.

Note that the solutions (A4) take into account all corollaries of the quantum regression theorem. Indeed, for the free parts of the operators $a_k(t)$ and $a_k^\dagger(t-\tau)$ we can derive the expressions

$$\langle A_k^{f\dagger}(t-\tau)A_k^f(t)\rangle=\int_{-\infty}^{\infty}d\omega\,|\kappa(\omega)|^2\times\exp\{-i\omega\tau\}\frac{\langle b_\omega^\dagger(0)b_\omega(0)\rangle}{\Gamma^2(\omega_k)+(\omega-\omega_k)^2}=n(k)\exp\{-i\omega_k\tau-\Gamma|\tau|\},$$

where $n(k)$ is the average number of photons in the kth mode of the microcavity.

Now, inserting (A4) into Eq. (2), we can eliminate $A_k^{f\dagger}(t)$ and $A_k^f(t)$ (by employing the Bogolyubov lemma[14]). Since

$$\langle a_k^\dagger(0)B(t)\rangle=n(\omega_k)\langle[B(t),a_k^\dagger(0)]\rangle,$$

$$\langle b_\omega^\dagger(0)B(t)\rangle=n(\omega)\langle[B(t),b_\omega^\dagger(0)]\rangle,$$

we can express the correlation function $\langle A_k^{f\dagger}(t)B(t)\rangle$ in terms of $a_k^{s\dagger}(t)$:

$$\langle A_k^{f\dagger}(t)B(t)\rangle = n(\omega_k)\langle [a_k^{s\dagger}(t), B(t)]\rangle,$$

where we have put $n(\omega)\approx n(\omega_k)$.

*)E-mail: enik@usm.md

[1] E. M. Purcell, Phys. Rev. **69**, 681 (1946).
[2] D. Kleppner, Phys. Rev. Lett. **47**, 233 (1981).
[3] G. Breit and E. Tetter, Astrophys. J. **91**, 215 (1940); R. Marrus and R. W. Schmieder, Phys. Lett. A **32**, 431 (1970); R. W. Schmieder and R. Marrus, Phys. Rev. Lett. **25**, 1245 (1970); R. Marrus and R. W. Schmieder, Phys. Rev. A **5**, 1160 (1972).
[4] B. Nikolaus, D. Z. Zhang, and P. E. Toshek, Phys. Rev. Lett. **47**, 171 (1981).
[5] M. Brune, J. M. Raimond, P. Goy, L. Davidovich, and S. Haroche, Phys. Rev. Lett. **59**, 1899 (1987); M. Brune, J. M. Raimond, and S. Haroche, Phys. Rev. A **35**, 154 (1987).
[6] W. Lange, G. S. Agarwal, and H. Walther, Phys. Rev. Lett. **76**, 3293 (1996).
[7] N. A. Enaki, Zh. Éksp. Teor. Fiz. **94**(10), 135 (1988) [Sov. Phys. JETP **67**, 2033 (1988)].
[8] Zh. Chen and H. Freedhoff, Phys. Rev. A **44**, 546 (1991); N. A. Enaki and M. A. Macovei, Phys. Rev. A **56**, 3274 (1997).
[9] N. A. Enaki and O. B. Perepelitsa, Opt. Spektrosk. **69**, 617 (1990) [Opt. Spectrosc. **69**, 367 (1990)].
[10] I. M. Beterov and P. B. Lerner, Usp. Fiz. Nauk **159**, 665 (1989) [Sov. Phys. Usp. **32**, 1084 (1989)].
[11] D. Meschede, H. Walther, and G. Müller, Phys. Rev. Lett. **54**, 551 (1985).
[12] G. Rempe, H. Walther, and N. Klein, Phys. Rev. Lett. **58**, 353 (1987).
[13] W. Lange and H. Walther, Phys. Rev. A **48**, 4551 (1993); G. S. Agarwal, W. Lange, and H. Walther, Phys. Rev. A **48**, 4555 (1993).
[14] N. N. Bogolyubov and N. N. Bogolyubov, Jr., Fiz. Élem. Chastits At. Yadra **11**, 245 (1980) [Sov. J. Part. Nucl. **11**, 93 (1980)].

Translated by Eugene Yankovsky

Radiative asymptotic behavior of stimulated Raman scattering

A. A. Zabolotskiĭ*)

Institute of Automation and Electrometry, Siberian Branch, Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 19 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 1168–1195 (April 1999)

This paper uses an integrable model to study an asymptotic solution describing the transformation of energy occurring in stimulated Raman scattering. The model allows for motion of populations and for the nonlinear Stark effect. Initial conditions leading to a radiative solution are discussed. The boundary conditions reflect the injection into the medium of high-power pulses of constant-amplitude pump and Stokes fields. It is shown that the radiative asymptotic behavior of this problem in the limit of weak medium excitation and in the limit of rapidly varying intense fields is determined by the kernels of Marchenko equations that are proportional to functions depending only on a self-similar variable. Analytic solutions are found for these cases. Detailed numerical calculations carried out for weak fields corroborate the analytic results. The role of the soliton part of the continuous spectrum of the problem is also studied. It is found that a high-power soliton of the Stokes field can be generated at the leading edge of a wave packet. © *1999 American Institute of Physics.* [S1063-7761(99)00204-8]

1. INTRODUCTION

Stimulated Raman scattering (SRS) of light has been studied for a long time (see, e.g., Ref. 1), but the interest in the problem is unflagging, which is due to the universality and relative simplicity of realizing this phenomenon in experiments. When high-power pump and Stokes fields are injected into the medium, packets of pulses are generated in the SRS process. The study of the nature and characteristics of such pulses is important for practical reasons.

An analysis of the SRS model often encounters analytical difficulties, which emerge when one wishes to describe the behavior of dense packets of pulses with many degrees of freedom. Some of these difficulties can be overcome by using models that are exactly solvable or are close to integrable. The most detailed information about the evolution of fields in nonlinear media can be extracted by the inverse scattering method (ISM).[2] As is known, the Maxwell–Bloch equations that describe SRS in one-dimensional media are integrable by ISM[3,4] for physical initial and boundary conditions. The evolution equations of the SRS model are similar to those of other physical models, for instance, the model of four-wave mixing in a medium with a Kerr nonlinearity, which in turn formally coincides with the ''uniaxial'' chiral model on the O_3 group, and the like.[5] In view of this, the analysis and the methods of solution used in the SRS model are also interesting from a theoretical standpoint.

In models with strong nonlinearities, such as the Maxwell–Bloch equations of one-particle interaction of a field and a two-level medium and the SRS mode, the radiative part of the solution can provide the main contribution to the interaction dynamics. The radiative solution describes, for instance, the quasi-self-similar asymptotic behavior of a long laser amplifier.[6] In the work of Gabitov, Manakov, Mikhaĭlov, Novokshenov, and Zakharov, the inverse scattering method is used to find related self-similar asymptotes of the SRS model[7,8] and the Maxwell–Bloch equations for media with nondegenerate[9,10] and degenerate[11] one-particle transitions. Gabitov *et al.*[9] studied the mixed boundary-value problem for the Maxwell–Bloch equations in a two-level medium for a one-particle transition.

Mathematically, the Chu–Scott model,[12] which was used in Refs. 8, 13, and 14 to describe SRS, is equivalent to the Maxwell–Bloch equations for a two-level single-particle laser amplifier.[15] The boundary conditions for the Chu–Scott model, corresponding to injection into the medium of pump and Stokes fields with constant amplitudes, were studied by Kaup and Menyuk.[13,14] However, in these papers no analytic solutions describing the asymptotic behavior were found.

The present paper is a study of SRS in a two-level medium. A totally integrable model is adopted. The model describes changes in level populations, pump depletion, and the nonlinear Stark effect. We examine the boundary conditions corresponding to injection into a sample end of wave packets of the pump and Stokes fields with arbitrary (but constant) amplitudes, and the physical initial conditions for the density matrix of the medium that lead to a contribution of the real spectrum into the asymptotic behavior. We also establish, for the general case, the Marchenko equations and the approximate asymptotic expression for the kernel corresponding to radiative asymptotic behavior. The explicit form of the radiative solution of the Marchenko equations will be found for the limit of weak medium excitation and for the limit of rapid energy exchange between strong fields and the medium. Computer simulation, done in the limit of weak medium excitation, is used to test the analytic results.

The analytic solutions found for the above boundary conditions and describing radiative asymptotic behavior is

also new for the Chu–Scott model[13] and hence for the mathematically identical model of a laser amplifier.[6] The proposed method of finding the radiative asymptotic behavior and the analytic results can applied to other models of nonlinear physics, such as the chiral field model and the model of the interaction of two polarized waves propagating in a two-level medium with a Kerr nonlinearity.[5,16]

Generally, when high-power pulses of the Stokes and pump fields are injected into the medium, in addition to accounting for the real spectrum one must account for the ''soliton'' part of the spectrum. We find that for an infinitely long steplike pulse, a part of the continuous spectrum may emerge in the ''soliton'' region of the spectral plane. For physical applications it is often important to establish the conditions required for the formation of a pulse with a high-power leading edge. An analysis of the dependence of the dynamics of a soliton packet on the initial conditions generated in such a problem makes it possible to find the conditions under which a soliton with the largest amplitude and the shortest duration forms at the leading edge of the packet.

The plan of the paper is as follows. Section 2 is devoted to the statement of the problem. In Sec. 3 we discuss the ISM equations and find an approximate asymptotic expression for the kernel of the Marchenko equations. In Sec. 4 we use the inverse scattering method to study the case of weak excitation of the medium. We find a radiative solution describing asymptotic solutions for arbitrary values of the pump- and Stokes-field amplitudes and the boundary of the sample. The analytic results are then compared with numerical calculations. Section 5 is devoted to the case of rapidly varying fields. We find the explicit form of the analytic solution describing radiative asymptotic behavior. In Sec. 6 we use the analytic results to explain the anomalies in the shape of Stokes-field pulses, anomalies observable in real experiments. In Sec. 7 we study the dynamics of a soliton packet that may be generated in the system. Finally, in the Appendix we establish the dependence of the constant-amplitude pump- and Stokes-fields on time and on slowly varying initial data.

2. BASIC EQUATIONS

We assume that the pump field and the Stokes field with amplitudes E_1 and E_2, respectively, propagate in a semi-infinite one-dimensional medium positioned along the x axis with a frequency-independent refractive index $n(\omega_j)=n_j$:

$$E(x,t)=\sum_{1,2}\left(\frac{\hbar\omega_j}{2cn_j}\right)^{1/2}\times[\mathbf{e}_jE_j(x,t)\exp\{i(k_jx-\omega_jt)\}+\text{c.c.}],$$

where $\mathbf{e}_j$ is the polarization vector, and ω_j and k_j are the carrier frequency and the wave vector, respectively. The condition for resonance in a two-level medium with a transition frequency ω_0 has the form $\omega_1-\omega_2=\omega_0+\nu_0$, where ν_0 is the frequency offset. Throughout this paper, with the exception of Sec. 7, it is assumed that $\nu_0=0$.

In the slow-envelope approximation, the Maxwell–Bloch equations describing stimulated Raman scattering are (see, e.g., Ref. 20)

$$\left(\partial_x+\frac{1}{v_1}\partial_t\right)E_1=-ib_1(N_3-N_0)E_1+i\kappa_0R^*E_2,$$

$$\left(\partial_x+\frac{1}{v_2}\partial_t\right)E_2=-ib_1(N_3-N_0)E_2+i\kappa_0RE_1,$$

$$\partial_tR=2i(b_1|E_1|^2+b_2|E_2|^2)R+i\kappa_0E_1E_2^*N_3,$$

$$\partial_tN_3=i\kappa_0E_1E_2^*R+\text{c.c.} \tag{1}$$

The phase velocities are equal: $v_1=v_2=v$. Here $v_i=c/n_i=\omega_i/k_i$, with c the speed of light; b_1, b_2, and κ_0 are the coefficients of cubic nonlinear susceptibility in the two-level medium expressed in terms of the physical constants of the medium, which can be found, for example, in Ref. 4; N_3 is the population difference of the levels involved in the transitions; N_0 is the number of atoms; R_0 is the polarizability of the medium; and x and t are the space and time coordinates, respectively.

Now we write the system of equations (1) in the form

$$\partial_TR_+=i[gR_+F_3+R_3F_+]-2i\nu_0R_+,$$

$$\partial_TR_3=-\partial_zF_3=\frac{i}{2}[R_+F_--R_-F_+],$$

$$\partial_zF_+=i[gF_+R_3+F_3R_+]. \tag{2}$$

Here

$$g=\frac{b_1-b_2}{\kappa_0},\quad z=\int_0^x\kappa_0N_0(s)\,ds,$$

$$N_0^2(x)=N_3^2+|R|^2,\quad T=\kappa_0\int_0^\tau I_1(y)\,dy,$$

$$\tau=t-\frac{z}{c},\quad I_1(T)=|E_1|^2+|E_2|^2,$$

$$F_3=\frac{|E_1|^2-|E_2|^2}{I_1},\quad R_3=\frac{N_3}{N_0},$$

$$R_+=\frac{R}{N_0}\exp\left[i(b_1+b_2)\int_0^\tau I_1(y)\,dy\right],$$

$$F_+=2\exp\left[i(b_1+b_2)\int_{-\infty}^\tau I_1(s)\,ds\right]\frac{E_1E_2^*}{I_1},$$

$$F_-=F_+^*,\quad R_-=R_+^*.$$

The systems of equations (2) has the integrals

$$|R_+|^2+R_3^2=1,\quad |F_+|^2+F_3^2=1. \tag{3}$$

The initial conditions that should be considered are those leading to the appearance of a contribution of the real spectrum to the asymptotic behavior. It is assumed that the initial excitation of the medium is described by the slowly varying functions of coordinates $R_3(z,0)$ and $R_+(z,0)$. The boundary conditions are fixed in the following way. Wave packets (of the pump field and the Stokes field) with constant amplitudes

$E_1(0,\tau)=E_1(0,0)$ and $E_2(0,\tau)=E_2(0,0)$ are introduced into the nonlinear medium at $z=0$. Numerical analysis conducted for $|g|<1$ has shown that the radiative solution for this model describes the transition of the system to the steady ground state. This state corresponds to the values of the fields and the medium polarizations at infinity ($z\to\infty$):

$$R_3(z,T)=-1, \quad R_+(z,T)=0,$$

$$F_3(z,T)=-1, \quad F_+(z,T)=0. \tag{4}$$

In Sec. 7 we study infinitely long pulses of the Stokes and pump fields and partial uniform excitation of the medium at the initial moment.

3. INVERSE SCATTERING METHOD

The inverse scattering method is basically used in the present investigation to study the radiative asymptotic behavior of the model (2) with the initial and boundary conditions specified above. The soliton and finite-band solutions of model (2) were built in Refs. 17, 5, and 18. The radiative asymptotic behavior of this model was studied in Ref. 19, in which the special case of $E_2(0,\tau)=0$ and of an exponentially small initial polarization,

$$\left|\ln\int_0^\infty |R_+(z,0)|\,dz\right| \gg 1,$$

was examined.

At $\nu_0=0$ the system of equations (2) can be written as the compatibility condition for the following two systems of linear equations:[3,4]

$$\partial_z\Phi=L\Phi=\begin{pmatrix} -i(\zeta-g/2)R_3 & (\zeta+\varphi_+)R_+ \\ -(\zeta+\varphi_-)R_- & i(\zeta-g/2)R_3 \end{pmatrix}\Phi, \tag{5}$$

$$\partial_T\Phi=A\Phi=\frac{1}{4\zeta}\begin{pmatrix} i(2\zeta g-1)F_3 & 2(\zeta+\varphi_+)F_+ \\ -2(\zeta+\varphi_-)F_- & i(2\zeta g-1)F_3 \end{pmatrix}\Phi, \tag{6}$$

where ζ is a spectral parameter, Φ is a two-component function, and $\varphi_+=-g/2\pm(i/2)(1-g^2)^{1/2}$.

We write the spectral problem (5) in the form

$$\partial_z\Phi=i\begin{pmatrix} -\lambda n & -(\lambda+\phi_+)\mu \\ -(\lambda+\phi_-)\bar{\mu} & \lambda n \end{pmatrix}\Phi, \tag{7}$$

where $n=R_3$, $\mu=iR_+$, $\bar{\mu}=-iR_-$, $\lambda=\zeta-g/2$, $\phi_\pm=\varphi_\pm+g/2$, and $\lambda=\zeta-g/2$.

Now we consider the case in which $g^2<1$. Under this condition the system tends to the ground state (4) as $z\to\infty$. The solutions of the system that correspond to the boundary condition $z=0$ and the asymptotic state (4) have the form $\chi_0^{-,+}=\exp(-i\lambda\sigma_3 z)$, with $z=0$ and $z\to\infty$, respectively. Let $\chi_0^{-,+}$ be the fundamental matrices of the Jost solutions ($\chi^{-,+}\to\chi_0^{-,+}$ as $z\to 0$ and $z\to\infty$, respectively). The spectral problem (7) meets the involution condition, i.e.,

$$\Phi(\lambda,z)=M\Phi(\lambda^*,z)^*M^{-1}, \tag{8}$$

where $M=\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. We define the scattering matrix $\mathscr{S}$ as follows:

$$\chi^-=\chi^+\mathscr{S}(\lambda,t), \quad \mathscr{S}=\begin{pmatrix} a & -b^* \\ b & a^* \end{pmatrix}, \quad \det\mathscr{S}=1. \tag{9}$$

The second column of the Jost matrix

$$\chi^{-,+}=\begin{pmatrix} \psi_1^{-,+} & \bar{\psi}_1^{-,+} \\ \psi_2^{-,+} & \bar{\psi}_2^{-,+} \end{pmatrix}$$

is analytic in the upper half-plane of λ. The functions $a(\lambda)$ and $\bar{b}(\lambda)$ are analytic in the upper half-plane, and $a(\lambda)$ has zeros λ_j there, which are the eigenvalues of the spectral problem (28). From (8) and (9) it follows that

$$\bar{\psi}_1^+=\frac{\bar{\psi}_1^-}{a}+\frac{b^*}{a}\psi_1^+, \tag{10}$$

$$\bar{\psi}_2^+=\frac{\bar{\psi}_2^-}{a}+\frac{b^*}{a}\psi_2^+. \tag{11}$$

We represent $\psi^+(z,\lambda)$ in the form

$$\psi^+(z,\lambda)=\Phi_0(z,\lambda)+\int_z^\infty ds \times\begin{pmatrix} \lambda K_1(z,s) & (\lambda+\phi_+)K_2(z,s) \\ -(\lambda+\phi_-)K_2^*(z,s) & \lambda K_1^*(z,s) \end{pmatrix} \times\Phi_0(s,\lambda). \tag{12}$$

Substituting (12) into (10), allowing for (8), and integrating with the weights

$$\int_{-\infty}^\infty \frac{\exp(-i\lambda s)\,(\lambda+\phi_-^*)}{2\pi(\lambda+\phi_+)}d\lambda, \quad \int_{-\infty}^\infty \frac{\exp(-i\lambda s)}{2\pi}d\lambda,$$

we obtain the Marchenko equations

$$\partial_z K_2(z,y)-i\phi_-K_2(z,y)+iF(z+y)+\int_z^\infty \partial_y F(y+s)K_1(z,s)\,ds=0, \tag{13}$$

$$\partial_z K_1^*(z,y)+\int_z^\infty K_2^*(z,s)[\partial_y F(y+s)-i\phi_+F(y+s)]\,ds=0, \tag{14}$$

where

$$F(z,T)=\int_{\mathscr{C}}\frac{b^*}{a}\frac{\exp\{-i\lambda z\}}{2\pi}d\lambda. \tag{15}$$

Note that $\rho=(b^*/a)(\lambda,T)$ is the scattering coefficient, which contains all information needed to find the radiative solution. The contour $\mathscr{C}$ consists of the real axis and passes above the poles in the upper complex half-plane. The soliton contribution is studied in Sec. 7. Up to that section we limit ourselves to studying only the radiative part of the spectrum, i.e., integration in (15) is carried out along the real axis.

The relationships that link the diagonal parts of the kernels $K_{1,2}(z,z,T)$ and the ''potentials'' $n(z,T)$ and $r(z,T)$ can be found from (6) and (12):

$$i\mu(z,T)K_2^*(z,z,T)=[1+iK_1(z,z,T)][n(z,T)-1], \tag{16}$$

$$[1+iK_1(z,z,T)]\mu^*(z,T)=-i[1+n(z,T)]K_2^*(z,z,T). \tag{17}$$

This yields

$$\mu(z,T)=\frac{2iK_2(z,z,T)U(z,T)}{|U(z,T)|^2+|K_2(z,z,T)|^2}, \tag{18}$$

$$n(z,T)=\frac{|U(z,T)|^2-|K_2(z,z,T)|^2}{|U(z,T)|^2+|K_2(z,z,T)|^2}, \tag{19}$$

where $U(z,T)=1+iK_1(z,z,T)$.

Solving the Marchenko equations requires calculating the kernel (15) with allowance for the time dependence of the scattering data (105) found in the Appendix. In this section we take into account only the contribution of the real continuous spectrum $\mathscr{C}_r$, which determines the radiative part of the solution of the problem. We replace λ with $\lambda-g/2$. If we use the expression (105), the kernel (15) assumes the form

$$F(z+y,T)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\lambda \times\exp\left[i\left(\frac{g}{2}-\lambda\right)(z+y)\right]\frac{r\exp(-2i\Omega T)+p}{c\exp(-2i\Omega T)+d}. \tag{20}$$

Here

$$\Omega(\lambda)=\pm\frac{1}{2}\sqrt{\left(\frac{1}{2\lambda}-q\right)^2+(1-g^2)|F_+^2(0,0)|},$$

$$r(\lambda)=A_{12}+\rho_0(i\Omega-A_{11}),$$

$$c(\lambda)=\rho_0A_{21}+i\Omega+A_{11},$$

$$p(\lambda)=-A_{12}+\rho_0(i\Omega+A_{11}),$$

$$d(\lambda)=-\rho_0A_{21}+i\Omega-A_{11},$$

with A_{ij} being the values of the components of the matrix $\|A\|$ on the right-hand side of Eq. (6) at $z=0$.

The denominator on the right-hand side of Eq. (20) vanishes at the points

$$-T\Omega(\lambda_n)=\pi\left(n+\frac{1}{2}\right)+i\kappa(\lambda_n),\quad \kappa(\lambda_n)=\frac{1}{2}\ln\frac{d}{c}(\lambda_n). \tag{21}$$

The sign of Ω is selected so that within the limits $F_+(z=0,T)=0$ and $F_3(z=0,T)=\pm1\ \mathrm{sgn}\,\Omega$ coincides with the sign of $F_3(z,T)\equiv F_3(0,T)$, since within these limits the time dependence can easily be found and has the simple form

$$\rho(\lambda,T)=\rho_0\exp\left[i\int_0^T F_3(0,T)\,dT\left(1-\frac{1}{2\lambda}\right)\right].$$

We find the asymptotic expression for the kernel $F(z,T)$ for large values of T. Estimating the integral by the saddle-point method, we can show that the main contribution to (20) is provided by the neighborhood of the point $\lambda_s\sim\sqrt{T}$. For large values of λ we have the expansion

$$\Omega(\lambda_n)\approx\varepsilon\left(\Omega_0+\frac{\alpha}{\lambda}\right)+O\left(\frac{\alpha_1}{\lambda}\right)^2, \tag{22}$$

where $\Omega_0=(1/2)\sqrt{(1-g^2)|F_+(0,0)|^2+g^2}$, $\varepsilon=\pm1$, $\alpha=-g/4\Omega_0$, and α_1 is a constant of order unity. For physical media, g can be either positive or negative. To be definite, we assume that g is positive. In this section we ignore the last term on the right-hand side of Eq. (22). Note that in the limits discussed below, $\alpha_1=0$.

The dependence of ρ_0 on λ is also found in the Appendix, where we show that

$$\rho_0=\frac{\lambda-i\phi_-^*}{\lambda}\frac{R_+(0,0)}{\delta+R_3(0,0)}\left[1+O\left(\frac{1}{\lambda}\right)\right]$$

for almost all slowly varying initial conditions. It can also be shown that in general, for λ large, the dependence of the coefficient of the exponentials on the right-hand side of Eq. (15) on the spectral parameter has the form

$$r=r_0+O\left(\frac{1}{\lambda}\right),\quad p=p_0+O\left(\frac{1}{\lambda}\right),$$

$$c=c_0+O\left(\frac{1}{\lambda}\right),\quad d=d_0+O\left(\frac{1}{\lambda}\right).$$

Now we find the kernel $F(z,T)$ with the λ-dependence of these coefficients ignored, i.e., we calculate the integral

$$F(z+y,T)=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\lambda\exp\left[i\left(\frac{g}{2}-\lambda\right)\times(z+y)\right]H(\lambda,T), \tag{23}$$

where

$$H(\lambda,T)=\frac{p_0+r_0\exp(-2i\Omega T)}{d_0+c_0\exp(-2i\Omega T)}.$$

Let $F_3(0,T)\equiv F_3(0,0)=\cos\beta_0=\varepsilon|\cos\beta_0|$. We call the interaction mode corresponding to the boundary conditions $\pi/2\geqslant\beta_0>0$ the *I*-mode, and the mode corresponding to the boundary conditions $\pi>\beta_0>\pi/2$ the *J*-mode. We show that the kernels corresponding to these interaction modes are proportional to the Bessel functions I_k and J_k $(k=0,1)$, respectively.

Calculating the residues on the right-hand side of F [Eq. (23)] at the poles (21),

$$\lambda_n=\frac{-\varepsilon\alpha T}{\pi n+\pi/2-i\kappa_0+\varepsilon\Omega_0T},\quad \kappa_0=\frac{1}{2}\ln\frac{d_0}{c_0},$$

we obtain a series in n. Multiplying the nth term of this series by the exponential with the exponent

$$2i\varepsilon\alpha T(\pi n+\pi/2-i\kappa+\varepsilon\Omega_0T)/\varepsilon\alpha T - i\pi-2\kappa_0-2i\varepsilon\Omega_0T,$$

we obtain

$$F(z+s,T)=-i\frac{r_0d_0-p_0c_0}{d_0^2}\frac{1}{2\pi i}\sum_n \exp\left[-\frac{i(z+s)}{q_n}+i2\alpha Tq_n+i\frac{g}{2}(z+y)-2i\varepsilon\Omega_0T\right], \quad (24)$$

where $q_n=\varepsilon(\pi n+\pi/2-i\kappa)/\alpha T+\Omega_0/\alpha$. For large T, the interval between q_n and q_{n+1} gets smaller, which justifies the transition from summation over q_n to integration with respect to q. For $\operatorname{sgn}\varepsilon=+1$, which corresponds to the *I*-mode, the kernel is

$$\begin{aligned}F(z+y,T)&=-\frac{g_\infty}{2\pi}\int_{-\infty}^{\infty}\exp\left[\frac{i(z+y)}{q}+2i\alpha Tq\right]\times\exp\left[-2iT\Omega_0+i\frac{g}{2}(z+y)\right]\frac{dq}{q^2}\\&=g_\infty\frac{4|\alpha|T}{\sqrt{8|\alpha|(z+y)T}}I_1(\sqrt{8|\alpha|(z+y)T})\times\exp\left[-2iT\Omega_0+i\frac{g}{2}(z+y)\right],\end{aligned} \quad (25)$$

where I_1 is a Bessel function, and $g_\infty=(r_0d_0-p_0c_0)/d_0^2$.

In the *J*-mode, the solution for $F_+(z,T)$ asymptotically $(z\to\infty)$ tends to zero. We select $\operatorname{sgn}\varepsilon=-1$, so that the solution matches the asymptotic solution in the linear limit, since a zero asymptote for $F_+(z,T)$ corresponds to $F_3|_{z\to\infty}=-1$. Replacing the signs of ε and q in the above formulas, we obtain a kernel that corresponds to the *J*-mode:

$$F(z+y,T)=g_\infty\frac{4|\alpha|T}{\eta}J_1(\eta)\exp\left[2iT\Omega_0+i\frac{g}{2}(z+y)\right], \quad (26)$$

where $J_1(\eta)$ is a Bessel function, and $\eta=\sqrt{8|\alpha|(z+y)T}$.

Note that the asymptotic behavior of the kernel $F(z,T)$ for large values of T and arbitrary (but constant) fields $E_{1,2}(z=0,T)=\text{const}$ is linearly dependent on the Bessel functions, as in the limiting cases: $F_3(0,T)\approx\pm1$.

The next step consists in solving the Marchenko equations (13) and (14) with the kernel $F(z+y,T)$. The solution can be found by iterations that use a series expansion in powers of $1/\eta$ for η large. The solution for $F_+(z,T)$ consists of a series of spikes with an amplitude decreasing as $z\to\infty$.

We show that solving the problem amounts to solving a first-order differential equation. It is known that the spectral problem (5) reduces to the Zakharov–Shabat spectral problem.[2] The corresponding gauge transformation has the form[3]

$$\Psi_z=D^{-1}\Phi,$$

$$D=[I\cos(\gamma/2)+i\sigma_3\sin(\gamma/2)][I\cos(v/2)+i\sigma_1\sin(v/2)][I\cos(\theta/2)+i\sigma_3\sin(\theta/2)], \quad (27)$$

where σ_i are the Pauli matrices, and we have introduced the notation

$$R_+=\exp\{\pm i\theta\}\sin v,\quad R_3=\cos v,\quad \gamma=\int_0^z\theta_z\cos v\,dz,$$

$$\theta_z=\partial_z\theta,\quad \gamma\to0,\quad z=0;\quad v,\ \theta_z\to0,\quad z\to\infty.$$

The transformation (27) reduces the spectral problem (5) to

$$\partial_z\Psi=\begin{pmatrix}-i\lambda & \frac{1}{2}V\\ -\frac{1}{2}V^* & i\lambda\end{pmatrix}\Psi, \quad (28)$$

where

$$V(z,T)=[(i\sqrt{1-g^2}-\theta_z)\sin v+iv_z]e^{i\gamma}. \quad (29)$$

For $1>g^2$, the corresponding Marchenko equations have the form[2]

$$K_1^{(1)}(z,y)+F^{(1)}(z+y)+\int_0^z F^{(1)}(y+s)K_2^{(1)}(z,s)\,ds=0, \quad (30)$$

$$K_2^{(1)*}(z,y)-\int_0^z K_1^{(1)*}(z,s)F^{(1)}(y+s)\,ds=0, \quad (31)$$

$$F^{(1)}(z)=\int_{\mathscr{C}}\frac{b^*}{a}\frac{\exp(-i\lambda z)}{2\pi}d\lambda. \quad (32)$$

The relationship between the ''potential'' V and the kernels $K_{1,2}$ has the form[2]

$$V(z,T)=4K_1^{(1)}(z,z,T), \quad (33)$$

$$\int_0^z|V(y,T)|^2\,dy=-4K_2^{(1)}(z,z,T). \quad (34)$$

The time dependence $\rho(T)$ can be bound by replacing matrix $\|A\|$ with matrix $\|A\|_g$ obtained from $\|A\|$ via the gauge transformation (27). For $F_{3,+}(0,T)\equiv F_{3,+}(0,0)$ and $R_{3,+}(z,0)\equiv R_{3,+}(0,0)$ (i.e., constants), the corresponding components of $\|A\|_g$ are independent of z and T, and the dependence of ρ on λ and time T is given by the general expressions (105) in the Appendix.

We can find the general form of the solution of the system of Marchenko equations (30) and (31) by following the ideas developed by Gabitov and Manakov.[11] Here it is possible to express $V(z,T)/T$ in terms of a function depending solely on a self-similar variable. Then, to reconstruct the ''potential'' $R_+(z,T)$, we must solve Eqs. (29) and (33), but there is no way in which this can be done analytically. At the same time, Eqs. (30) and (31) also emerge in the physically interesting limit of weak medium excitation, a limit in which (29) becomes trivial and corresponds to the formal equalities

$$V(z,T)\equiv R_+(z,T),\quad R_3(z,T)\equiv-1. \quad (35)$$

In this limit, the gauge transformation (27) becomes an identity transformation.

4. WEAK-FIELD LIMIT

The Chu–Scott model[12] can be obtained from Eqs. (2) in the weak-field limit ($|\lambda| \ll |\phi_{\pm}|$). If we ignore excitation [see Eq. (35)] and the Stark effect ($g=0$), we obtain the Chu–Scott model:

$$\partial_T V = F_+ ,$$

$$\partial_z F_+ = F_3 V, \tag{36}$$

$$\partial_z F_3 = -\tfrac{1}{2}(F_+^* V + F_+ V^*). \tag{37}$$

The functions and the variables are the same as in (2). The Chu–Scott model can be represented by the compatibility condition for the linear systems (28) and the system

$$\partial_T \Phi = \frac{1}{4\lambda}\begin{pmatrix} iF_3 & -iF_+ \\ -iF_- & -iF_3 \end{pmatrix}\Phi. \tag{38}$$

The Marchenko equations coincide with (30) and (31). The time dependence of the scattering coefficient is given by the general expression (105) in the Appendix, where $\|A\|$ is the matrix on the right-hand side of Eq. (38) at $z=0$. For the Chu–Scott model we have the exact relationships

$$\Omega = \frac{\varepsilon}{4\lambda}, \quad \varepsilon = \frac{F_3(0,0)}{|F_3(0,0)|}, \quad \alpha = \frac{1}{4}, \quad \alpha_1 = 0. \tag{39}$$

The coefficients on the right-hand side of Eq. (20) are of the form

$$r^{(1)}(\lambda) = \frac{1}{4\lambda}[i\rho_0^{(1)}(\varepsilon - F_3(0,0)) + F_+(0,0)] \times\left[1 + O\left(\frac{1}{\lambda}\right)\right] \approx \frac{r_0^{(1)}}{4\lambda}, \tag{40}$$

$$p^{(1)}(\lambda) = \frac{1}{4\lambda}[i\rho_0^{(1)}(\varepsilon + F_3(0,0)) - F_+(0,0)] \times\left[1 + O\left(\frac{1}{\lambda}\right)\right] \approx \frac{p_0^{(1)}}{4\lambda}, \tag{41}$$

$$c^{(1)}(\lambda) = \frac{1}{4\lambda}[i(\varepsilon + F_3(0,0)) - \rho_0^{(1)}F_-(0,0)] \times\left[1 + O\left(\frac{1}{\lambda}\right)\right] \approx \frac{c_0^{(1)}}{4\lambda}, \tag{42}$$

$$d^{(1)}(\lambda) = \frac{1}{4\lambda}[i(\varepsilon + F_3(0,0)) + \rho_0^{(1)}F_-(0,0)] \times\left[1 + O\left(\frac{1}{\lambda}\right)\right] \approx \frac{d_0^{(1)}}{4\lambda}. \tag{43}$$

We ignore the term $O(1/\lambda)$ in square brackets, i.e., in (40)–(43) $\rho_0^{(1)}$, $r_0^{(1)}$, $p_0^{(1)}$, $c_0^{(1)}$, and $d_0^{(1)}$ are independent of λ. Allowing for (39), we find that the kernel $F^{(1)}(x+y,T)$ is given by Eqs. (25) and (26). For the *J*-mode we have

$$F^{(1)}(z+y,T) = \frac{g_\infty^{(1)}}{\sqrt{2(z+y)T}} J_1(\sqrt{2(z+y)T}), \tag{44}$$

where $g_\infty^{(1)} = (r_0^{(1)}d_0^{(1)} - p_0^{(1)}c_0^{(1)})/(d_0^{(1)})^2$, and J_1 is a Bessel function. For the *I*-mode the kernel differs from (44) in that J_1 is replaced by I_1. When $R_+(z,0)=0$, we can easily show that $\rho_0(\lambda)=0$, with

$$g_\infty^{(1)}|_{\rho_0=0} = \frac{2|\sin\beta_0|}{(1+|\cos\beta_0|)^2}.$$

To find the radiative solution of the Marchenko equations (30) and (31), we use the Gegenbauer addition formula, following the ideas developed in Ref. 11. Examining the *J*-mode, we introduce the variables $\xi = \sqrt{2zT}$ and $\zeta = \sqrt{2yT}$ and write the function $F^{(1)}(z+y,T)$ in the form $F^{(1)}(z+y,T) = T\mathscr{F}^{(1)}(\sqrt{\xi^2+\zeta^2})$. Next we expand the function $\mathscr{F}^{(1)}(\sqrt{\xi^2+\zeta^2})$ in Bessel functions:

$$\mathscr{F}^{(1)}(\sqrt{\xi^2+\zeta^2}) = \frac{g_\infty^{(1)}}{\sqrt{\xi^2+\zeta^2}} J_1(\sqrt{\xi^2+\zeta^2}) = g_\infty^{(1)}\frac{2}{\zeta\xi}\sum_{k=1}^{\infty}(-1)^{k-1}(2k-1) \times J_{2k-1}(\zeta)J_{2k-1}(\xi). \tag{45}$$

Here we have used the properties of the Gegenbauer polynomial C_l^k (see, e.g., Ref. 21):

$$C_l^{2k-1}(0) = 0, \quad C_l^{2k}(0) = (-1)^k\frac{(k+l)!}{l!\,k!}.$$

We introduce $\mathscr{K}_{1,2}^{(1)}(z,y) = T^{-1}K_{1,2}^{(1)}(z,y,T)$ and expand these functions in the Bessel functions J_{2k-1}:

$$\mathscr{K}_1^{(1)}(z,y) = \mathscr{K}_1^{(1)}(\xi,\zeta) = \frac{2}{\xi\zeta}\sum_{k=1}^{\infty}\mathscr{R}_k^{(1)}(\xi)J_{2k-1}(\zeta), \tag{46}$$

$$\mathscr{K}_2^{(1)}(z,y) = \mathscr{K}_2^{(1)}(\xi,\zeta) = \frac{2}{\xi\zeta}\sum_{k=1}^{\infty}\mathscr{P}_k^{(1)}(\xi)J_{2k-1}(\zeta). \tag{47}$$

We take advantage of the biorthogonality of the Bessel functions and Neumann's Bessel functions Y_k,

$$\int_{\mathscr{C}} J_m(\zeta)Y_k(\zeta)\,d\zeta = a_k\delta_{km},$$

$$a_0 = 2\pi i, \quad a_k = \pi i, \quad k>0, \tag{48}$$

where the integral is taken along a contour $\mathscr{C}$ that encloses the center of the complex plane. Multiplying the Marchenko equations by Y_{2k-1} and integrating along $\mathscr{C}$, we find that

$$\mathscr{R}_l^{(1)}(\xi) = \sum_{n=1}^{\infty}\mathscr{O}_{ln}^{(1)}\mathscr{P}_n^{(1)}(\xi) + \mathscr{F}_k^{(1)}(\xi), \tag{49}$$

$$\mathscr{P}_l^{(1)}(\xi) = -\sum_{n=1}^{\infty}\mathscr{O}_{ln}^{(1)}\mathscr{R}_n^{(1)}(\xi), \tag{50}$$

where

$$\mathscr{F}_k^{(1)} = (-1)^{k-1}2g_\infty^{(1)}(2k-1)J_{2k-1}(\xi),$$

$$\mathscr{O}_{kl}^{(1)}=(-1)^{k-1}4g_{\infty}^{(1)}(2k-1)\mathscr{W}_{kl}^{(1)},$$

$$\mathscr{W}_{kl}^{(1)}(s)=\int_0^s J_{2k-1}(\sigma)J_{2l-1}(\sigma)\sigma^{-1}\,d\sigma. \tag{51}$$

For the *I*-mode we have a similar algebraic system of equations in which the Bessel functions J_k are replaced by I_k. The integrals in (51) reduce to tabulated integrals, e.g.,

$$\int I_{2k-1}(x)I_{2j-1}(x)x^{-1}\,dx = 2^{1-2j-2k}x^{-2+2j+2k}F[(j+k-1,j+k-1/2),(2j,2k,2k+2j-1),x^2]/[(j+k-1)\Gamma(2j)\Gamma(2k)],$$

where F is the hypergeometric function and Γ is the gamma function. We find the solution to the algebraic system of equations (49) and (50) using Cramer's formula

$$\mathscr{P}_k^{(1)}=\frac{\det\|M_k^{(1)}\|}{\det\|M^{(1)}\|}, \tag{52}$$

where

$$\|M^{(1)}\|=\delta_{ij}+\mathscr{O}_{ik}^{(1)}\mathscr{O}_{kj}^{(1)};$$

$\|M_k^{(1)}\|$ differs from $\|M^{(1)}\|$ in that the kth column is replaced by the vector $\mathscr{F}_k^{(1)}$. We have

$$\frac{1}{4\xi}\frac{d}{d\xi}\ln\det\|M^{(1)}\|=\frac{1}{\det\|M^{(1)}\|}\sum_{k=1}^{\infty}\det\|M_k^{(1)}\|J_{2k-1}. \tag{53}$$

Comparing (52) with (53) and using the rule of differentiation of determinants, we find the formal solution for $\mathscr{K}_2^{(1)}$ [Eq. (47)]:

$$\sum_{n=1}^{\infty}\mathscr{P}_k^{(1)}(\xi)J_{2k-1}(\xi)=\mathscr{K}_2^{(1)}(\xi,\zeta)=\frac{1}{4\xi}\frac{\partial}{\partial\xi}\ln\det(\delta_{kl}+\|\mathscr{O}^{(1)}\|\|\mathscr{O}^{(1)*}\|). \tag{54}$$

To find $\mathscr{K}_1^{(1)}(\xi,\zeta)$, we write (49) and (50) in vector form, bearing in mind that $\|\mathscr{O}^{(1)}\|$ is a real matrix:

$$(I+\|\mathscr{O}^{(1)}\|^2)\mathscr{R}^{(1)}=\mathscr{F}^{(1)}. \tag{55}$$

We write the solution of Eq. (55) also in vector form:

$$\mathscr{R}^{(1)}=\frac{1}{2i}[(I+i\|\mathscr{Q}^{(1)}\|)^{-1}\mathscr{F}^{(1)}+(I-i\|\mathscr{Q}^{(1)}\|)^{-1}\mathscr{F}^{(1)}]. \tag{56}$$

The final result is

$$\mathscr{K}_1^{(1)}(\xi,\zeta)=\frac{1}{4\xi}\frac{\partial}{\partial\xi}\ln\frac{\det[I+i\|\mathscr{O}^{(1)}\|]}{\det[I-i\|\mathscr{O}^{(1)}\|]}. \tag{57}$$

Instead of (33) and (34) we have

$$V(\zeta,T)=4T\mathscr{K}_1^{(1)}(\zeta,\zeta), \tag{58}$$

$$\int_0^{\zeta}|V(\eta)|^2\,d\eta=-4T\mathscr{K}_2^{(1)}(\zeta,\zeta). \tag{59}$$

Equation (58) implies that the field $V(z,\tau)$ is proportional to a function that depends solely on the self-similar variable $\zeta=\sqrt{2zT}$. The solution that describes the dynamics of the field $V(\zeta,T)$ has the form

$$V(\zeta,T)=\frac{T}{\xi}\frac{\partial}{\partial\xi}\ln\frac{\det[I+i\|\mathscr{O}^{(1)}\|]}{\det[I-i\|\mathscr{O}^{(1)}\|]}. \tag{60}$$

Note that for $g^2<1$ the solution of Eq. (29) with the left-hand side (60) makes it possible to find the radiative asymptotic behavior for the general case of arbitrary excitation of the medium and with the Stark effect taken into account. Obviously, in this case the solution is not self-similar, although it is determined by the self-similar function $\mathscr{K}_1(\zeta,\zeta)$.

The radiative part of the solution of the system of equations (2) for $\pi/2>\beta_0>0$ consists of two parts: an increasing part, proportional (for small T) to the Bessel function I_1, anddamped oscillations (see Figs. 1 and 2). For $\pi>\beta_0>\pi/2$, the solution consists of damped oscillations (Fig. 3). The corresponding kernel $\mathscr{F}^{(1)}(\eta)$ is proportional to the Bessel function $J_1(\eta)$.

In Refs. 6 and 9, where the Marchenko equations were studied for a one-particle laser amplifier, it is shown that for an initially almost totally inverted medium the solution of the equations also exhibits a similar dependence on a self-similar variable. In these papers, the researchers used initial conditions that correspond to a small deviation from total inversion of the medium and to field fluctuations. A more general solution is found in the present paper for a situation corresponding, in the case of a laser amplifier, to an arbitrary degree of inversion of the medium.

We assume that

$$-\log[|\beta_0|^{-1}]\gg 1, \quad -\log[|\rho_0|^{-1}]\gg 1. \tag{61}$$

In a laser amplifier, these conditions [Eq. (61)] correspond to almost total initial inversion. The small Bloch angle β_0 describes quantum fluctuations of the polarizability of the medium.[9] If $\beta_0=0$, conversion of the medium occurs at $\rho_0\neq 0$ (see Ref. 6). Similarly, for the adopted model with $\rho_0=0$, an initial Stokes field $E_2(0,\tau)$ is required to initiate the energy conversion process, and the shape of this field determines the leading edge of the asymptotic solution. Strictly speaking, the classical model is inapplicable when $\beta_0=\rho_0=0$, since it does not allow for quantum fluctuations of vacuum and medium. However, when the number of photons is large, the classical model provides a satisfactory description of the dynamics of the fields.

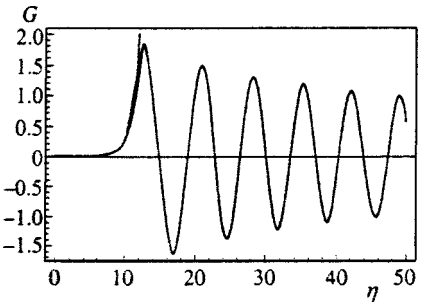

FIG. 1. The *I*-mode of interaction. The dependence of $G(\eta)=(\eta/4T)V(\eta,T)$ (see Eq. (36)) on the self-similar variable $\eta=\sqrt{4zT}$ depicted in this figure was found numerically for $\beta_0=10^{-4}$ and $\rho_0=0$. The graph of the linear solution found for small T, i.e., the function $g_0 I_1(\eta)$, is also shown. It lies above the graph of the numerical solution of Eq. (62).

For real initial values of the fields and zero frequency offset, the self-similar solution of the Chu–Scott model is described by the equation

$$\mathscr{B}''_{\eta\eta}(\eta)+\frac{1}{\eta}\mathscr{B}'_{\eta}(\eta)=\sin\mathscr{B}(\eta), \tag{62}$$

where

$$V(z,T)=\frac{4T}{\eta}\mathscr{B}'_{\eta}(\eta), \quad \eta=\sqrt{4zT}.$$

In the Marchenko equations (13) and (14) we can ignore the integrals for small values of η. When η is large, the nonlinear terms can be ignored only for small enough coefficients $|g_\infty|$ in the *J*-mode. In the first case the linear solution of the Marchenko equations describes the leading edge of the solution corresponding to the *I*- and *J*-modes. We examine the integral (15) in the limit of small T such that the denominator on the right- hand side of Eq. (20) does not vanish. For T and $\sigma=1/\lambda$ small, the factor $\exp(2i\Omega T)$ in the denominator can be replaced by 1. We write (20) in the form

$$F_0^{(1)}(z+s,T)=-\frac{\sin\beta_0}{4\pi}\int_{\mathscr{C}}\frac{d\sigma}{\sigma^2}[1-\exp(2iT\sigma)]\times\exp\left[-\frac{i(z+s)}{\sigma}\right]. \tag{63}$$

Integrating along $\mathscr{C}$ in (63) in the positive sense about the singularity $\sigma=0$, we obtain

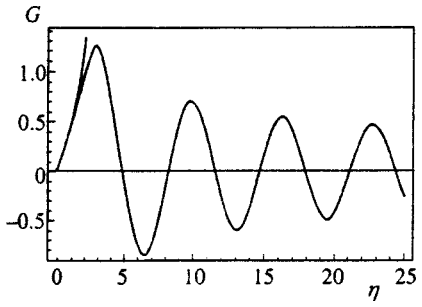

FIG. 2. The same as in Fig. 1, but for $\beta_0=0.25\pi$.

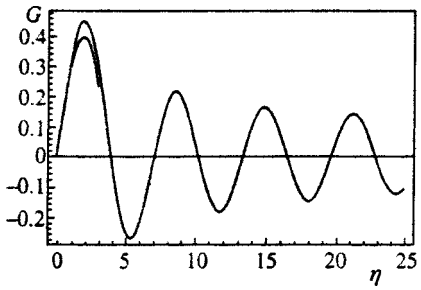

FIG. 3. The *J*-mode of interaction. The same as in Fig. 1 for $\beta_0=0.75\pi$. The graph for the linear solution $g_0 J_1(\eta)$ is also shown. It lies below the graph of the numerical solution.

$$F_0^{(1)}(z+s,T)=-\frac{\sin\beta_0}{8\pi}\int_{\mathscr{C}_+}\frac{d\sigma}{\sigma^2}\times\exp\left[-\frac{i}{\sigma}(z+s)+2iT\sigma\right]=g_0^{(1)}\frac{T}{\sqrt{\zeta^2+\xi^2}}I_1(\sqrt{\zeta^2+\xi^2})=T\mathscr{F}_0^{(1)}, \tag{64}$$

where $\zeta=\sqrt{2Tz}$, $\xi=\sqrt{2Ts}$, I_1 is a Bessel function, and $g_0^{(1)}=|\sin\beta_0|$. The solution for the *J*-mode can be found from (64) by replacing I_1 with J_1.

The other linear limit corresponds to the *J*-mode and to small $|g_\infty|$, i.e., $\pi-\beta_0\ll\pi$. The corresponding kernel has the form [see (26)]

$$F^{(1)}(z+y,T)=\frac{g_\infty^{(1)}T}{\eta}J_1(\eta), \quad \eta=\sqrt{2(z+y)T}.$$

The solution (26) describes damped oscillations of the field $V(z,T)$ about the stable state (4).

In the first linear limit (small T and η), the solution of the Marchenko equations (13) and (14) is obvious:

$$\mathscr{K}_1^{(1)}(\zeta,\zeta)=\mathscr{F}_0^{(1)}(\zeta),$$
$$V(z,T)=4K_1^{(1)}(\zeta,\zeta,T)=4Tg_0^{(1)}\mathscr{F}_0^{(1)}(\zeta)=\frac{4Tg_0^{(1)}}{\eta}I_1(\eta), \tag{65}$$

where $\eta=\sqrt{2}\,\zeta=\sqrt{4zT}$. Accordingly, for the *J*-mode we must replace I_1 with J_1. When η is small, the solution (65) is valid for all β_0 (see Figs. 1–3). In the second linear limit, the solution can be found by introducing the formal substitutions $I_1\to J_1$ and $g_0^{(1)}\to g_\infty^{(1)}$ into (65). The Marchenko equations do not specify the sign of the field, so we find it by matching the resulting solution and the linear solution.

Figure 4 compares the numerical investigation of the self-similar asymptotic solution of the Chu–Scott model and the solution (65) expressed in terms of $\mathscr{F}^{(1)}$ [see Eq. (32)]. Numerical analysis has shown that at $\beta_0=0.9\pi$ the maximum deviation of the numerical solution from the analytic one at the first vertex is less than 1%. For $\beta_0=0.8\pi$ the deviation is no larger than 3%. And $\beta_0=0.7\pi$ the deviation

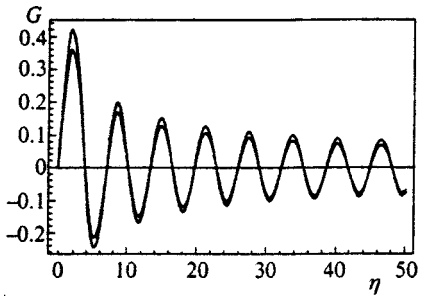

FIG. 4. Comparison of the numerical results with the analytic results of the Chu–Scott model in the *J*-mode. The dependence of $G(\eta)=(\eta/4T)V(\eta,T)$ on the self-similar variable $\eta=\sqrt{4zT}$ depicted in this figure was found numerically for $\beta_0=0.8\pi$ and $\rho_0=0$. The graph of the linear solution $g_\infty J_1(\eta)$, which has a smaller amplitude of oscillations, is also shown.

is no larger than 8%. If we continue to reduce β_0, the error gets larger and the linear approximation breaks down.

The numerical results show that the analytic solution found in the second linear approximation can provide a satisfactory description of the radiative solution from $\beta_0=\pi$ to $\beta_0\approx 0.75\pi$, with g_∞ varying from 0 to roughly 1.1.

Computer simulation of the nonlinear region shows that the theory correctly describes the radiative solution at large and small *T*. The amplitude of oscillations of the radiative part of the solution depends on β_0, while the general form of the radiative part of the solution remains the same for all angles β_0, which corroborates the self-similar nature of the asymptotic behavior of SRS for the initial and boundary conditions specified in this section.

5. LIMIT OF RAPIDLY VARYING HIGH-POWER FIELDS

Modern laser technology makes it possible to use high-power laser fields in studies of nonlinear processes. If the number of photons of the pump and Stokes fields passing through unit volume element of the medium is larger than the active atoms in this element, there is rapid energy exchange between fields and medium. When the fields are strong, the *z*-derivative of the kernel $K_{1,2}$ is proportional to λ and can reach values greater than unity. In real physical media, $|\phi_+|$ varies between 0.15–1.5. We assume that $|\lambda|\gg|\phi_+|$ in the strong-field limit. In this limit we can use the inverse scattering method, eliminating $\phi_\pm$ from the equations of Sec. 3. For instance, the solutions of the simplified spectral problem ($g=1$) are related by

$$\overline{\psi_1^+}=\frac{\overline{\psi_1^-}}{a}+\frac{b^*}{a}\psi_1^+. \tag{66}$$

The solution of the reduced spectral problem can be written

$$\chi^+(z,\lambda)=\Phi_0^+(z,\lambda)-\int_z^\infty ds\begin{pmatrix}\lambda K_1^{(0)}(z,s) & \lambda K_2^{(0)}(z,s)\\ -\lambda\overline{K}_2^{(0)}(z,s) & \lambda\overline{K}_1^{(0)}(z,s)\end{pmatrix}\times\Phi_0^+(s,\lambda). \tag{67}$$

The Marchenko equations can be derived by integrating (66) with the weight

$$\int_{-\infty}^{\infty}\frac{\exp(-i\lambda s)}{2\pi\lambda}d\lambda.$$

The final result is

$$K_1^{(0)}(z,y)+F^{(0)}(z+y)+\int_0^z F^{(0)}(y+s)K_2^{(0)}(z,s)\,ds=0, \tag{68}$$

$$K_2^{(0)*}(z,y)-\int_0^z K_1^{(0)*}(z,s)F^{(0)}(y+s)\,ds=0, \tag{69}$$

where

$$F^{(0)}(z+y)=\int_{\mathscr{C}}\frac{\rho(\lambda,z)}{2\pi\lambda}\exp\left[-i\left(\lambda-\frac{1}{2}\right)(z+y)\right]d\lambda. \tag{70}$$

The ρ vs. *T* dependence is also given by Eq. (105) in the Appendix. The kernel (70) is calculated in the same way as in Sec. 3, with the following exact relationships employed in the calculations:

$$\Omega=\frac{\varepsilon}{4\lambda},\quad \varepsilon=\frac{F_3(0,0)}{|F_3(0,0)|},\quad \Omega_0=-\frac{1}{2},$$

$$\alpha=\frac{1}{4},\quad g=1. \tag{71}$$

We can show (as we did earlier) that the coefficients of the exponents on the right-hand side of Eq. (70) allows an expansion $r=r_0+O(1/\lambda)+\cdots$, where r_0 is independent of λ. Substituting the expression for ρ_0 found in the Appendix and repeating the procedure discussed in Sec. 3, we find the kernel corresponding to the *I*-mode:

$$F^{(0)}(z+y,T)=q_\infty^{(0)}I_0(\sqrt{2(z+y)T})\times\exp\left[\frac{i}{2}(z+y)-iT\right]. \tag{72}$$

For the kernel corresponding to the *J*-mode we have

$$F^{(0)}(z+y,T)=g_\infty^{(0)}J_0(\sqrt{2(z+y)T})\times\exp\left[\frac{i}{2}(z+y)+iT\right]. \tag{73}$$

Here for g_∞ we have (with allowance for the relationship $\rho_0=iR_+(0,0)/[\delta+R_3(0,0)]$ found in the Appendix)

$$g_\infty^{(0)}=\frac{4\varepsilon[\sin\beta_0(\delta+\cos v_0)^2-\cos\beta_0\sin^2 v_0+i\sin\beta_0(\delta+\cos v_0)]}{(\delta+\cos v_0)(\varepsilon+\cos\beta_0)-\sin\beta_0\sin v_0}, \tag{74}$$

$\delta=\operatorname{sgn} R_3(0,0)$, $R_3(0,0)=\cos v_0$, $R_+(0,0)=\sin v_0$.

Next we limit ourselves to the *J*-mode. We expand $F^{(0)}(z+y,T)$ of (73) in Bessel functions:

$$\begin{aligned}F^{(0)}(\sqrt{\xi^2+\zeta^2},z)&=g_\infty^{(0)}J_0(\sqrt{\xi^2+\zeta^2})\\&\quad\times\exp\left[i\frac{1}{4T}(\xi^2+\zeta^2)+iT\right]\\&=g_\infty^{(0)}\sum_{k=0}^{\infty}(-1)^k 2kJ_{2k}(\zeta)J_{2k}(\xi)\\&\quad\times\exp\left[i\frac{1}{4T}(\xi^2+\zeta^2)+iT\right],\end{aligned} \tag{75}$$

where $\zeta=\sqrt{2zT}$, and $\xi=\sqrt{2yT}$. Then we expand $\mathscr{K}_{1,2}^{(0)}=T^{-1}K_{1,2}^{(0)}$ in Bessel functions:

$$\begin{aligned}\mathscr{K}_1^{(0)}(z,y,T)&=\mathscr{K}_1^{(0)}(\xi,\zeta,T)\\&=\sum_{k=1}^{\infty}\mathscr{R}_k^{(0)}(\xi)J_{2k}(\zeta)\exp\left[i\frac{1}{4T}(\xi^2+\zeta^2)+iT\right],\end{aligned} \tag{76}$$

$$\begin{aligned}\mathscr{K}_2^{(0)}(z,y,T)&=\mathscr{K}_2^{(0)}(\xi,\zeta,T)\\&=\sum_{k=1}^{\infty}\mathscr{P}_k^{(0)}(\xi)J_{2k}(\zeta)\exp\left[i\frac{1}{4T}(\xi^2-\zeta^2)+iT\right].\end{aligned} \tag{77}$$

Using the biorthogonality property (48) and repeating the procedure described in Sec. 3, we obtain from (68) and (69) the algebraic equations

$$\mathscr{R}_l^{(0)}(\xi)=\sum_{n=1}^{\infty}\mathscr{O}_{ln}^{(0)}\mathscr{P}_n^{(0)}(\xi)+\mathscr{G}_l^{(0)}(\xi), \tag{78}$$

$$\mathscr{P}_l^{(0)}(\xi)=-\sum_{n=1}^{\infty}\mathscr{O}_{ln}^{(0)}\mathscr{R}_n^{(0)}(\xi), \tag{79}$$

where

$$\mathscr{O}_{kl}^{(0)}=g_\infty^{(0)}(-1)^k 2k\mathscr{W}_{kl}^{(0)},$$

$$\mathscr{G}_k^{(0)}(\xi)=g_\infty^{(0)}(-1)^k 2kJ_{2k}(\xi).$$

The integrals

$$\mathscr{W}_{kl}^{(0)}(s)=\int_0^s J_{2k}(\sigma)J_{2l}(\sigma)\sigma\,d\sigma \tag{80}$$

have been tabulated:

$$\begin{aligned}&\int I_{2k}(x)I_{2j}(x)x\,dx\\&\quad=2^{-1-2j-2k}x^{2+2j+2k}F[(j+k+1/2,1+j+k,1+j\\&\quad+k),(2j+1,2+j+k,1+2k,1+2j+2k),x^2]/\\&\quad[(j+k+1)\Gamma(2j+1)\Gamma(2k+1)],\end{aligned} \tag{81}$$

$$\begin{aligned}&\int J_{2k}(x)J_{2j}(x)x\,dx\\&\quad=2^{-1-2j-2k}x^{2+2j+2k}F[(j+k+1/2,1+j+k,1+j\\&\quad+k),(2j+1,2+j+k,1+2k,1+2j+2k),-x^2]/\\&\quad[(j+k+1)\Gamma(2j+1)\Gamma(2k+1)],\end{aligned}$$

where F is the hypergeometric function and Γ is the gamma function. Equations (78) and (79) can be solved in the same way as in Sec. 4. Using Cramer's formula (52), where $\|M^{(0)}\|=\delta_{ij}+\mathscr{O}_{ik}^{(0)}\mathscr{O}_{kj}^{(0)}$ ($\|M_k^{(0)}\|$ differs from $\|M^{(0)}\|$ in that the kth column is replaced by vector $\mathscr{G}_k^{(0)}$), we find

$$\mathscr{K}_2^{(0)}(\xi,\xi)=\frac{1}{\partial\xi}\frac{\partial}{\partial\xi}\ln\det(\delta_{kl}+\|\mathscr{O}^{(0)}\|\|\mathscr{O}^{(0)*}\|). \tag{82}$$

We can find the kernel $\mathscr{K}_1^{(0)}(\xi,\xi)$ in the same way as we did earlier, using (78) and (79) for the real matrix $\|\mathscr{O}^{(0)}\|$. Writing (78) and (79) in vector form,

$$(I+\|\mathscr{O}^{(0)}\|^2)\cdot\mathscr{R}^{(0)}=\mathscr{G}^{(0)}, \tag{83}$$

we have

$$\begin{aligned}\mathscr{K}_1^{(0)}(\xi,\xi)&=\frac{1}{2i\xi}\frac{\partial}{\partial\xi}\left\{\ln\frac{\det[I+i\|\mathscr{O}^{(0)}\|]}{\det[I-i\|\mathscr{O}^{(0)}\|]}\right\}\\&\quad\times\exp\left(\frac{i}{2T}\xi^2+iT\right).\end{aligned} \tag{84}$$

We introduce the functions

$$\mathscr{Q}_\pm^{(0)}=\frac{1}{\xi}\frac{\partial}{\partial\xi}\ln\det(I\pm i\|\mathscr{O}^{(0)}\|) \tag{85}$$

and write the above solution in the form

$$\mathscr{K}_2^{(0)}(\xi,\xi)=\frac{1}{2}(\mathscr{Q}_+^{(0)}+\mathscr{Q}_-^{(0)}), \tag{86}$$

$$\mathscr{K}_1^{(0)}(\xi,\xi)=\frac{1}{2i}(\mathscr{Q}_+^{(0)}-\mathscr{Q}_-^{(0)})\exp\left(\frac{i}{2T}\xi^2+iT\right). \tag{87}$$

Using (18) and (19), we find the solution:

$$R_+(z,T)=\frac{2\,\mathrm{Re}\,\mathscr{Q}_+^{(0)}[1+i\,\mathrm{Im}\,\mathscr{Q}_+^{(0)}\exp\{(i/2T)\xi^2+iT\}]}{1+(1/2)[\mathscr{Q}_+^{(0)2}+\mathscr{Q}_-^{(0)2}]-2\,\mathrm{Im}\,\mathscr{Q}_+^{(0)}\sin\{(1/2T)\xi^2+T\}}, \tag{88}$$

$$R_3(z,T)=\frac{1-\mathscr{Q}_+^{(0)}\mathscr{Q}_-^{(0)}-2\,\mathrm{Im}\,\mathscr{Q}_+^{(0)}\sin\{(i/2T)\xi^2+T\}}{1+(1/2)[\mathscr{Q}_+^{(0)2}+\mathscr{Q}_-^{(0)2}]-2\,\mathrm{Im}\,\mathscr{Q}_+^{(0)}\sin\{(1/2T)\xi^2+T\}}. \tag{89}$$

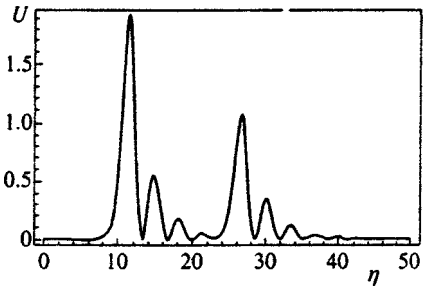

FIG. 5. The $U(\eta)=|(\eta/4T)V(\eta,T)|^2$ vs. $\eta=\sqrt{4zT}$ dependence. The anomalies in the shape of the Stokes pulse resulting from the mixing of the contributions of spontaneous and stimulated Raman scattering depicted in this figure were found for $\beta_0=0.01\pi$ and $\rho_0=10^{-4}$. The finiteness of the ''transverse'' relaxation time T_2 was taken into account. In the units used in Eqs. (2), $T_2=3$. The contributions of spontaneous and stimulated Raman scattering to the shape of the Stokes field were calculated individually. The figure depicts the superposition of the fields.

6. APPLYING THE RESULTS

We show that our results can be used to explain the anomalies in the shape of the Stokes field observed in Raman scattering experiments[22] within the context of the Chu–Scott model. The fields at the medium's boundary $z=0$ and the initial polarization of the medium $\{R_+(z,0),R_3(z,0)\}$ contribute to the kernel $F^{(1)}$. When ρ_0 is small, the scattering coefficient can be written in the form of a sum:

$$\rho(\lambda,T)\simeq\frac{A_{12}}{2i\Omega}[\exp(-2i\Omega T)-1]+\rho_0\exp(-2i\Omega T)+O(\rho_0^2,A_{12}^2,\rho_0A_{12}). \tag{90}$$

Here the first term on the right-hand side is associated with stimulated Raman scattering, and the second with spontaneous Raman scattering.

In the Chu–Scott model, $F_+(0,0)=\sin\beta_0\approx\beta_0$, so that we have

$$\rho(\lambda,T)\simeq\frac{-i\beta_0}{2\varepsilon}\left[\exp\left[-\frac{iT\varepsilon}{2\lambda}\right]-1\right]-\frac{iR_+(0,0)}{2\lambda\delta}\times\exp\left[-\frac{iT\varepsilon}{2\lambda}\right]. \tag{91}$$

Here $R_+(0,0)$ describes the quantum fluctuations of the medium and initiates spontaneous Raman scattering. Equation (91) implies that spontaneous and stimulated Raman scatterings are described by different types of kernel $F^{(1)}$. Indeed, the results of Sec. 4 suggest that the first and second terms on the right-hand side of Eq. (91) lead to expressions for the kernel that are proportional to the Bessel functions I_1 or J_1 and I_0 or J_0, respectively. When the ratio $|F_+(0,T)/R_+(z,0)|$ is much greater or much less than unity, the shape of the Stokes pulse differs substantially from that found, e.g., in Refs. 7 and 14 and depicted in Figs. 1–3.

It appears that formula (91) makes it possible to explain for the first time the anomalies in the shape of the Stokes-field pulse that are observed in cooperative Raman scattering in hydrogen vapor.[22] In these experiments, a high-power pump-field pulse was injected into the medium, and then the shape of the Stokes field was investigated. The Stokes pulse was initiated by fluctuations in the field and medium. After repeating this experiment many times, Rautian *et al.*[22] found that the shape of the generated Stokes pulse varies considerably from one experiment to the next. For instance, the amplitude of the first spike in the generated packet can be either greater or less than the amplitude of the next spike. In approximately one trial out of ten, a high-power pulse was observed at the trailing (decaying) edge of the Stokes field. To the author's knowledge, no meaningful description of this phenomenon exists in the literature.

Figures 5 and 6 depict the results of numerical calculations using the Chu–Scott model for various values of the ratio $F_+(0,T)/R_+(z,0)$. The numerical results show that the theory developed in the present paper can explain these anomalies in the shape of the Stokes field.

7. SOLITON ASYMPTOTIC BEHAVIOR

We study soliton generation at the leading edge of the Stokes field, considered to be an infinitely long step. We use the symmetry of the system of equations (2) under the permutations

$$F_3\leftrightarrow R_3,\quad F_\pm\leftrightarrow R_\pm,\quad z\leftrightarrow T. \tag{92}$$

With allowance for frequency offset and (92), the spectral problem assumes the form

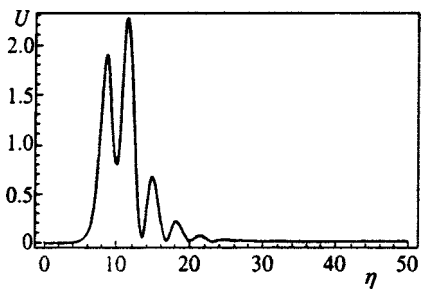

FIG. 6. The same as in Fig. 5, but for $\beta_0=10^{-3}\,\pi$ and $\rho_0=0.01$.

$$\partial_T\Phi=\begin{pmatrix} -i\lambda F_3 & (\lambda+\phi_+)F_+ \\ -(\lambda+\phi_-)F_- & i\lambda F_3 \end{pmatrix} \times\Phi+\begin{pmatrix} iv_0 & 0 \\ 0 & -iv_0 \end{pmatrix}\Phi, \tag{93}$$

where $\phi_\pm=\pm(1/2)\sqrt{1-g^2}$. All results remain valid, but their physical interpretation changes. To study the dynamics of a soliton packet we must (a) find the solution of the spectral problem (93) for a long steplike Stokes-field pulse propagating against the background of an infinitely long pump-field pulse, i.e.,

$$\begin{aligned} &F_+(0,T)=0, && T\geqslant T_0,\\ &F_+(0,T)=A_0\neq 0, && T<T_0, \end{aligned} \tag{94}$$

and (b) find the dependence of the scattering coefficient on z by using the linear system

$$\partial_z\Phi=\frac{1}{4\lambda+2g}\begin{pmatrix} i(2\lambda g-1+g^2)R_3 & 2(\lambda+\phi_+)R_+ \\ -2(\lambda+\phi_-)R_- & i(2\lambda g-1+g^2)R_3 \end{pmatrix}\Phi. \tag{95}$$

We assume that initially the medium is partially inverted, $R_+(z,0)=R_+(0,0)\neq 0$. The z-dependence of the scattering data is given by formulas in the Appendix with allowance for (92). There we also show that if the T-dependence of $F_+(0,T)$ is represented by an infinitely long step, the part of the continuum that lies in the upper half-plane may emerge. This spectrum is associated with generation of a soliton packet.

The effect of solitons on the asymptotic behavior in the case in which $E_{1,2}(0,T)$ are of finite length can be ignored if long time intervals are considered. This is due to the difference that exists between the group velocities of the radiative and soliton solutions. Indeed, suppose that the fields $E_{1,2}$ cross within some finite time interval $[0,T_0]$. The radiative solution describes the evolution of the system toward a stable state. If the initial condition $F_+(z,0)$ leads to the emergence of a pole in the upper half of the complex plane, a soliton solution will emerge and propagate against the background of the ground state. The time of soliton existence is limited by the soliton lifetime in the interval $[0,T_0]$. For the case of stimulated Raman scattering this fact was noted by Menyuk.[23]

The study of soliton asymptotic behavior has practical meaning only for large values of T_0. For physical applications it is important to establish the conditions under which the Stokes-field soliton, which has the highest amplitude, possesses the largest group velocity and propagates against the background of the stable ground state. Under these conditions, a packet of solitons forms with the highest-energy soliton at the leading edge.

We now find the conditions for establishing such a regime in the model (2) of stimulated Raman scattering. Important information about the soliton characteristics can be extracted from (15). We examine the spectrum associated with an infinitely long step (see Fig. 7). Contributions to the integral (15) are provided by F_c (the integral along the real axis), F_1 (the integral along the part of $\mathscr{C}_1$ lying in the upper half plane with the exception of the vertex λ_0), and F_0 (the integral around the vertex):

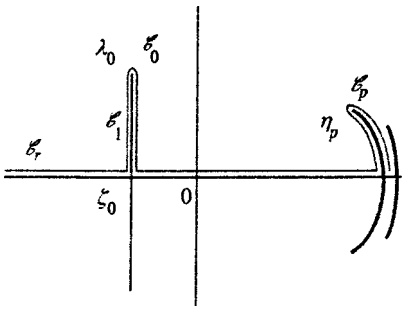

FIG. 7. The continuous spectrum associated with an infinitely long rectangular pulse $T_0\to\infty$. The spectrum incorporates the entire real axis $\mathscr{C}_r$ and $\mathscr{C}_s=\mathscr{C}_0\cup\mathscr{C}_1$, the ''soliton part'' of the continuous spectrum, where $\mathscr{C}_0$ is the neighborhood of the vertex λ_0. Here $\mathscr{C}_p$ is the part of the spectrum associated with the finite-band wave, and η_p is the point of this spectrum with the maximum imaginary part.

$$F(T+T',z)=F_c(T+T',z)+F_1(T+T',z)+F_0(T+T',z).$$

We examine the phase $\Theta=i[-(\lambda-g)(T+T')+\Omega z]$ in (15), where $\Omega(\lambda)$ is determined by the values of $R_{3,+}(T,z)\equiv R_{3,+}(T,0)$ at $T=0$. The group velocity Υ of the soliton packet can be found by assuming that $\operatorname{Im}\Theta=0$. In terms of the physical variables x and t we have $\Upsilon=c/n(\omega_1)\gamma(1+\gamma)^{-1}$, where $\gamma=\operatorname{Im}\Omega/\operatorname{Im}\lambda>0$ $(\operatorname{Im}\Omega(\lambda)>0)$.

Analysis of the phase $\Theta(z,T)$ and the group velocity Υ shows that the a faster soliton, propagating against the background of the ground state, has a smaller amplitude if $F_+(0,T)$ is a rectangular pulse [Eq. (94)]. We show that if the medium is not initially in the ground state, quite a different soliton generation regime is possible.

We analyze $\Theta(\lambda)$ for $\lambda\in\mathscr{C}_s$ in the neighborhood of the vertex λ_0. Since the highest and narrowest soliton corresponds to the vertex λ_0 of the soliton spectrum, it would be interesting to establish the conditions under which this soliton has the highest group velocity. For an arbitrary point λ in the spectrum, on a monotonic continuous curve in the soliton part of the spectrum, this condition is

$$\frac{d}{d\operatorname{Im}\lambda}\frac{\operatorname{Im}\Omega(\lambda)}{\operatorname{Im}\lambda}<0, \quad \operatorname{Im}\Omega(\lambda)>0. \tag{96}$$

Actually, this condition means that a soliton associated with a point λ such that $\operatorname{Im}\lambda=\eta<\eta_0=\operatorname{Im}\lambda_0$ has a group velocity that is lower than that of the soliton associated with the point λ_0.

If the Stokes and pump fields at $z=0$ are finite-band solutions of the problem, the associated spectrum may consist of a finite set of arcs that are symmetric about the real axis. We select the arc $\mathscr{C}_p$ with its vertex at the point η_p with the largest imaginary part (see Fig. 7). We can show that for the soliton associated with this vertex to have the maximum group velocity, the radius of curvature of the arc must be less than the distance from the origin of the complex plane to the arc's vertex.

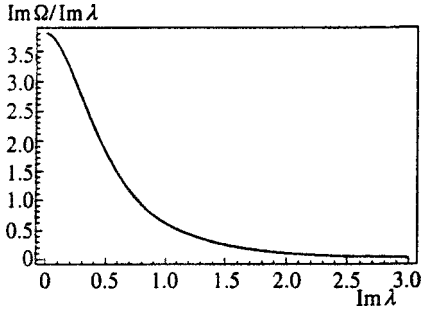

FIG. 8. $\mathrm{Im}\Omega(\lambda)/\mathrm{Im}\lambda$ vs. $\mathrm{Im}\lambda$ for $g=0.5$ and $\zeta_0=-0.3$.

The condition (96) is not satisfied when we are dealing with a steplike pulse of the fields and the initial state of the medium is the ground state. At the same time, this condition may be valid for a steplike pulse if initially the medium was partially inverted, i.e., if the medium was not initially in the ground state. Figure 8 depicts the dependence of $\mathrm{Im}\,\Omega(\lambda)$ on $\mathrm{Im}\lambda$ for $\lambda\in\mathscr{C}_s$, which dependence is possible, for example, for the following values of the parameters of the problem: $g>-0.4$ and $\nu_0 F_+(0,0)<0$. For such a dependence the condition (96) is satisfied.

We show that if the inequalities in (96) are valid, the Stokes-field soliton with the highest amplitude splits off from the wave packet. To prove this, we make the substitutions $T=\gamma_0 z+\vartheta$ and $T'=\gamma_0 z+\sigma$. Now we introduce the small parameter $\varepsilon=\eta(\gamma_0-\gamma)$ into (96), with $\gamma_0=\mathrm{Im}\,\Omega(\lambda_0)/\mathrm{Im}(\lambda_0)$. The parameter is positive if condition (96) is satisfied. Expanding in powers of ε, we find that

$$
\begin{aligned}
&F_0(T+T',z)\\
&=\frac{1}{2\pi}\int_{\eta}^{\eta_0} h(\lambda)\exp[-i(\eta-g)(T+T')+i2\Theta z]\,d\eta\\
&=\frac{1}{2\pi}\int_{\eta}^{\eta_0} h(\lambda)\exp\{-i(\zeta_0-g)(\vartheta+\sigma+2\gamma_0 z)\\
&+i2\mathrm{Re}\Theta(\lambda_0)z+\eta_0(\vartheta+\sigma)-\varepsilon[(\eta_1-\varepsilon\eta_2+\cdots)\\
&\times(\vartheta+\sigma)+z(1+im)]\}\,d\eta,
\end{aligned}
\tag{97}
$$

with $\lambda_0=\zeta_0+i\eta_0$.

The specific form of $h(\lambda)$ is unimportant for subsequent estimates and we omit it here. In (97) we used the expansion

$$
\eta=\eta_0+\varepsilon\eta_1-\varepsilon^2\eta_2+\cdots,\quad m=\mathrm{Re}\frac{d\Omega(\lambda)}{d\eta}.
$$

It can be shown that the expansion of $h(\eta)$ in powers of ε has the form

$$
h(\eta)=\sqrt{\varepsilon}\,(h_0+\varepsilon h_1+\varepsilon^2 h_1+\cdots).
$$

The expansion (97) implies that the soliton associated with the point λ_0 splits off from the wave packet. The overlap of the soliton and the remaining part of the wave packet decreases with increasing z, since $\eta_1>0$ and $\varepsilon z>0$. If condition (96) is satisfied, the distance from the leading soliton to the remaining part of the wave packet can be shown to increase as $\log z$.

This result can be expanded to the entire ''soliton'' part of the spectrum, $\mathscr{C}_s$. We show that when (96) is satisfied, the contribution of the entire soliton part to the leading soliton decreases with increasing z.

We find the contribution of the part of the soliton spectrum $\mathscr{C}_1$ not incorporating the neighborhood $\mathscr{C}_0$ of the vertex λ_0 (see Fig. 7) by estimating the integral F_1, which is calculated along $\mathscr{C}_1$ in the upper half-plane:

$$
\begin{aligned}
F_1(T+T',z)=&\frac{1}{2\pi}\int_{\mathscr{C}_1} h(\lambda)\\
&\times\exp[-i\lambda(T+T')+i2\Omega z]\,d\lambda.
\end{aligned}
\tag{98}
$$

To estimate the integral in (98), we write the imaginary part of the phase factor in the form

$$
\begin{aligned}
&\mathrm{Im}[\Theta(z,T,\lambda)]\big|_{T=\gamma_0 z,\ T'=\gamma_0 z}=z\,\mathrm{Im}[\tilde{\Theta}(\lambda)]\\
&=2z\,\mathrm{Im}\left[\frac{\lambda\,\mathrm{Im}\,\Omega(\lambda_0)}{\eta_0}-\Omega(\lambda)\right]
\end{aligned}
$$

where $\lambda\in\mathscr{C}_1$. Replacing integration over λ by integration over $\tilde{\Theta}(\lambda)$ and integrating by parts, we find that

$$
\begin{aligned}
&F_1(T+T',z)\\
&=\frac{h(\lambda_0)\exp[iz\,\mathrm{Re}\tilde{\Theta}(\lambda_0)]-h(\zeta_0)\exp[iz\tilde{\Theta}(\zeta_0)]}{iz\tilde{\Theta}'(\zeta_0)}\\
&\times\left[1+O\left(\frac{1}{z}\right)\right],
\end{aligned}
\tag{99}
$$

where $\zeta_0=\mathrm{Re}\lambda_0$. The estimate (99) implies that the contribution of the spectrum $\mathscr{C}_1$ to the shape of the leading soliton decreases with increasing distance in proportion to $1/z$, i.e., the distance from the leading soliton to the remaining part of the packet increases as $\log z$.

In the present paper we have established that in the event of stimulated Raman scattering, the adopted initial and boundary conditions (see the end of Sec. 2) lead to formation of radiative and soliton packets of pulses. When the conditions (96) and (99) are satisfied, a soliton with the maximum amplitude and group velocity appears at the leading edge of the soliton packet.

The results of this section are also of interest in substantiating the utility of Whitham's heuristic method.[2] This method is used to describe the development of modulation instability in nonlinear media. Using the single-phase solution is the most common approach to describing the transformation of a steplike pulse into a soliton packet. According to this method, the slow variation of the parameters of a periodic wave describes the transformation of a plane wave into a soliton packet. Since the development of modulation instability is due to long-wavelength excitations, the trailing edge of the packet should is by a quasiharmonic wave, while at the leading edge there is a maximum-amplitude soliton. For the SRS model discussed in this paper, this condition is met when initially there is partial inversion, $R_+(z,0)\sim O(1)$. At the same time, initial inversion leads to a radiative solution, and this solution cannot be described by the Whitham approach.

Note that an initial solution in the form of a steplike pulse also leads to the formation of a similar spectrum, i.e., a spectrum consisting of the real axis and a straight segment lying in the soliton part, in models that allow for a Lax representation of the Ablowitz–Kruskal–Newell–Segur type.[2] In particular, the condition (96) is satisfied for a (modified) Korteweg–de Vries equation. For such equations, a rectangular pulse splits into a soliton packet and a maximum-amplitude soliton at the leading edge. For a (modified) Schrödinger equation, the inequalities in (96) become equalities. The condition (96) is satisfied under special initial conditions, in which the soliton spectrum $\mathscr{C}_s$ is in the region $\mathrm{Re}\,\lambda<0$, and the distance from the point in the spectrum to the imaginary axis decreases monotonically with decreasing $\mathrm{Im}\,\lambda>0$ (Ref. 24).

An analysis of several integrable models with strong nonlinearities, such as the Maxwell–Bloch equations for a two-level medium,[3,4] and the four-wave mixing model,[5,16] has shown that initial and boundary conditions leading to the formation of a high-power soliton at the leading edge of the wave packet are also possible. This requires conditions similar to partial initial inversion. Under this condition one must take into account the contribution of the radiative solution, which can dictate the type of asymptotic solution. The results and approaches of the present investigation can be used for such nonlinear optics models.

This work was supported by grants from the Russian Fund for Fundamental Research [Grant No. 98-02-17904) and Deutsche Forschungsgemeinschaft (Grant No. 426 RUS 113/89/0(R,S)].

APPENDIX: TIME DEPENDENCE OF THE SCATTERING DATA

We begin with a finite interval $z\in[0,L_f]$. We assume that $F_{3,+}(0,T)$ is constant and that $F_+(L_f,T)$ is an arbitrary function of T. The dependence of the scattering data on T can be found by solving the linear system (6). To find this dependence, we write the solutions of both systems, (5) and (6), in the form

$$\Psi=\chi^+\Phi^+=\chi^-\Phi^-, \tag{100}$$

where $\chi^{\pm}$ are Jost functions. These functions are solutions of the system (5) but do not satisfy (6). The functions Φ^- and Φ^+ [solutions of (6) at $z=0$ and $z=L_f$, respectively] are such that

$$\partial_T\Phi^{\mp}=A\Phi^{\mp}. \tag{101}$$

Comparing Eqs. (9) and (100), we find that the matrix $\mathscr{S}$ satisfies the equation

$$\partial_T\mathscr{S}(z)=-\mathscr{S}A(z=0)+A(z-L_f)\mathscr{S}. \tag{102}$$

The components A_{11}, A_{22}, and A_{21} of the matrix $\|A\|$ at $z=L_f$ do not contribute to the T-dependence of the scattering data. The contribution of the component A_{12} at $z=L_f$ can be ignored. This was demonstrated by Kaup and Menyuk,[3,13] who found that the values of the matrix $\|A(z=L_f,T)\|$ determine the dynamics outside the ''physical'' region, i.e., for $z>L_f$. The interested reader is referred to those papers. For an infinite interval $L_f=\infty$ the system reaches the ground state as $z\to\infty$. Here $A_{12}\to\infty$ in the problem considered, i.e., the fields at infinity do not contribute to the T-dependence of the spectral parameter. To find the time dependence of the scattering data it is enough to restrict attention to the values of the matrix $\|A\|$ at $z=0$. The solution (102) for the coefficients a and b, to within a common factor, takes the form

$$a=[(i\Omega+A_{11})\,\mathrm{e}^{-i\Omega T}+(i\Omega-A_{11})\,\mathrm{e}^{i\Omega T}]\times a_0-b_0^*A_{21}(\mathrm{e}^{-i\Omega T}-\mathrm{e}^{i\Omega T}), \tag{103}$$

$$b^*=-a_0A_{12}(\mathrm{e}^{-i\Omega T}-\mathrm{e}^{i\Omega T})+b_0^*[(i\Omega-A_{11})\,\mathrm{e}^{-i\Omega T}+(i\Omega+A_{11})\,\mathrm{e}^{i\Omega T}]. \tag{104}$$

Here the components of $\|A\|$ at $z=0$ are such that $A_{11}=-A_{22}$ and $\Omega^2=-A_{11}^2-A_{12}A_{21}$.

Thus, the scattering coefficient ρ has the form

$$\rho(T)=\frac{b^*}{a}=-\frac{A_{12}(\mathrm{e}^{-i\Omega T}-\mathrm{e}^{i\Omega T})-\rho_0[(i\Omega-A_{11})\mathrm{e}^{-i\Omega T}+(i\Omega+A_{11})\mathrm{e}^{-i\Omega T}]}{(i\Omega+A_{11})\mathrm{e}^{-i\Omega T}+(i\Omega-A_{11})\mathrm{e}^{i\Omega T}-\rho_0A_{21}(\mathrm{e}^{-i\Omega T}-\mathrm{e}^{i\Omega T})}, \tag{105}$$

where

$$\rho_0=\rho(z=0,\lambda)=\frac{b_0^*}{a_0}(\lambda).$$

We wish to find the scattering coefficient ρ_0 under the assumption that the scale of variation of the functions $R_{+,3}(z,0)$ is much less than λ. The solution of the spectral problem (5) has the form

$$f_1(z)=f_1(0)e^{-i\lambda Z}+\frac{\lambda+\phi_+}{\lambda}\times\int_0^z\frac{R_+(u,0)}{R_3(u,0)}f_2(u)e^{-i\lambda(Z-U)}\,du, \tag{106}$$

$$f_2(z)=f_2(0)e^{i\lambda Z}-\frac{\lambda+\phi_-}{\lambda}\times\int_0^z\frac{R_-(u,0)}{R_3(u,0)}f_1(u)e^{i\lambda(Z-U)}\,du, \tag{107}$$

where

$$Z=\int_0^z R_3(u,0)\,du, \quad U=\int_0^u R_3(u,0)\,du.$$

We solve the system (106) and (107) iteratively, assuming that in the lowest order the solution is

$$f_1(z)\simeq f_1(0)\exp(-i\lambda\,\delta z), \quad f_2(z)\simeq f_2(0)\exp(i\lambda\,\delta z),$$

where $\delta=\pm 1=\operatorname{sgn} R_3(0,0)$. Integration by parts yields

$$\begin{aligned} f_1(z)=f_1(0)e^{-i\lambda Z}-\frac{\lambda+\phi_+}{i\lambda}\Bigg[&\frac{R_+(0,0)}{\delta+R_3(0,0)}f_2(0)e^{-i\lambda Z}\\ &-\frac{R_+(z,0)}{\delta+R_3(z,0)}f_2(z)\Bigg]-\frac{\lambda+\phi_+}{i\lambda}\\ &\times\int_0^z\frac{\partial}{\partial u}\left[\frac{R_+(u,0)}{\delta+R_3(u,0)}f_2(u)e^{-i\lambda U}\right]\\ &\times e^{-i\lambda(Z-U-\delta u)}\,du, \end{aligned} \tag{108}$$

$$\begin{aligned} f_2(z)=f_2(0)e^{i\lambda Z}-\frac{\lambda+\phi_-}{i\lambda}\Bigg[&\frac{R_-(0,0)}{\delta+R_3(0,0)}f_1(0)e^{-i\lambda Z}\\ &-\frac{R_-(z,0)}{R_3(z,0)}f_1(z)\Bigg]-\frac{\lambda+\phi_-}{2i\lambda}\\ &\times\int_0^z\frac{\partial}{\partial u}\left[\frac{R_-(u,0)}{\delta+R_3(u,0)}f_1(u)e^{i\lambda U}\right]e^{i\lambda(Z-U-\delta u)}\,du. \end{aligned} \tag{109}$$

Continuing to integrate by parts, we find the asymptotic expansion in powers of λ^{-1}. Here it is assumed that $R_+(u)/[\delta+R_3(u)]$, and that all its derivatives exist. This condition is satisfied for all physically justified initial conditions. Using the asymptotic expansion, we can find the appropriate expansion for the scattering data required by the given problem. Since the main contribution to the radiative solution is provided by large λ (see Sec. 3), we can limit consideration to the first terms in the expansion. Assuming that $f_1(0)=1$, $f_2(0)=0$, and $R_+(z,0)=0$, $z>l_0$, we find that

$$a_0\simeq 1+\frac{\lambda+\phi_+}{i\lambda}\frac{R_+(l_0,0)}{\delta+R_3(l_0,0)}f_2(l_0)e^{-i\lambda l_0\delta}\simeq 1, \tag{110}$$

$$\begin{aligned} b_0&\simeq-\frac{\lambda+\phi_-}{i\lambda}\frac{R_-(0,0)}{\delta+R_3(0,0)}\\ &\quad+\frac{\lambda+\phi_-}{i\lambda}\frac{R_-(l_0,0)}{\delta R_3(l_0,0)}f_1(Z_0)e^{-i\lambda l_0\delta}\\ &\simeq-\frac{\lambda+\phi_-}{i\lambda}\frac{R_-(0,0)}{\delta+R_3(0,0)}. \end{aligned} \tag{111}$$

Similarly, assuming that $f_1(0)=0$ and $f_2(0)=1$, we find that for the second column of the $\mathscr{S}$ matrix,

$$a_0\simeq 1, \quad b_0^*\simeq+i\frac{\lambda+\phi_-^*}{\lambda}\frac{R_+(0,0)}{\delta+R_3(0,0)}. \tag{112}$$

We now examine the boundary and initial conditions studied in Sec. 7. For an infinitely long step pulse of the Stokes field we find the associated spectrum of the problem. We investigate the case of a semi-infinite medium occupying the half-interval $[0,+\infty)$ into which the finite rectangular pulse (94) is injected.

When A_0 is constant, the solution of (5) can easily be found. For physical applications it is of interest to study the initial conditions corresponding to a finite frequency offset, since allowance for the frequency offset $\nu_0\neq 0$ results in substantial modifications of the spectrum of the problem.

Following a well-known procedure (see, e.g., Ref. 2), we can find the solution of (28) for the ''potential'' (94). The coefficient a_0 has the form

$$a_0=e^{2i\zeta T_0}\left[\cos(2\zeta T_0)+\frac{2\lambda}{\zeta}\sin(2\zeta T_0)\right], \tag{113}$$

where

$$\zeta=\sqrt{(\lambda-\nu_0\sqrt{1-A_0^2})^2+\tfrac{1}{4}(1-g^2+4\nu_0^2)|A_0^2|}\,.$$

The zeros $a_0(\zeta_n)=0$, which specify the spectrum of the problem, can be found from the equality

$$2\zeta_n T_0=-\arctan\frac{\zeta_n}{2\lambda}+n\pi+\frac{\pi}{2}. \tag{114}$$

An analysis of (114) shows that for T_0 finite there is a finite set of isolated poles; the poles in the upper half of the complex λ plane are associated with soliton solutions. The number of poles increases with T_0, while the distance between poles decreases in proportion to $1/T_0$. In the limit $T_0\to\infty$, the zeros ζ_n fill the entire real axis. For $1>g^2-4\nu_0^2$, a part of the spectrum may lie on a line parallel to the imaginary axis. This part is described by the segment $[\zeta_0-i\eta_0,\zeta_0+i\eta_0]$, where $\zeta_0=\nu_0\sqrt{1-A_0^2}$, and $\eta_0=(1/2)|A_0|\sqrt{1-g^2+4\nu_0^2}$ (see Fig. 7). The half of this segment in the upper half plane is associated with the soliton part of the spectrum, since it is obtained as a result of the merger of an infinite number of poles lying in that part of the plane.

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

[1] F. Reintjes, *Nonlinear Optical Parametric Processes in Liquids and Gases*, Academic Press, New York (1984).
[2] S. V. Manakov, S. P. Novikov, L. P. Pitaevskiĭ, and V. E. Zakharov, *Theory of Solitons: The Inverse Scattering Method*, Consultants Bureau, New York (1984).
[3] D. J. Kaup, Physica D **6**, 143 (1983).
[4] H. Steudel, Physica D **6**, 155 (1983).
[5] V. E. Zakharov and A. V. Mikhaĭlov, JETP Lett. **45**, 349 (1987).
[6] S. V. Manakov, Zh. Éksp. Teor. Fiz. **83**, 68 (1982) [Sov. Phys. JETP **56**, 37 (1982)].
[7] I. R. Gabitov, V. E. Zakharov, and A. V. Mikhaĭlov, JETP Lett. **37**, 279 (1983).
[8] I. R. Gabitov, V. E. Zakharov, and A. V. Mikhaĭlov, Zh. Éksp. Teor. Fiz. **86**, 1204 (1984) [Sov. Phys. JETP **59**, 703 (1984)].
[9] I. R. Gabitov, V. E. Zakharov, and A. V. Mikhaĭlov, Teor. Mat. Fiz. **63**, 11 (1985).
[10] S. V. Manakov and V. Yu. Novokshenov, Teor. Mat. Fiz. **69**, 40 (1986).
[11] I. R. Gabitov and S. V. Manakov, Phys. Rev. Lett. **50**, 495 (1983).
[12] F. Y. F. Chu and A. C. Scott, Phys. Rev. A **12**, 2060 (1975).
[13] D. J. Kaup and C. R. Menyuk, Phys. Rev. A **42**, 1712 (1990).

[14] C. R. Menyuk, Phys. Rev. A **47**, 2235 (1993).
[15] H. Steudel, Quantum Opt. **2**, 387 (1990).
[16] A. A. Zabolotskii, Physica D **40**, 283 (1989).
[17] R. Meinel, Opt. Commun. **49**, 224 (1984).
[18] I. V. Cherednik, Teor. Mat. Fiz. **47**, 225 (1981).
[19] A. A. Zabolotskiĭ, Zh. Éksp. Teor. Fiz. **93**, 84 (1987) [Sov. Phys. JETP **66**, 47 (1987)].
[20] H. Steudel, Exp. Tech. Phys. (Berlin) **20**, 409 (1972).
[21] *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun (eds.), National Bureau of Standards Applied Mathematics Series 55, Washington, D.C. (1964).
[22] S. G. Rautian, V. P. Safonov, and B. M. Chernobrod, Izv. Akad. Nauk SSSR, Ser. Fiz. **50**, 1513 (1986).
[23] C. R. Menyuk, Phys. Rev. Lett. **62**, 2937 (1989).
[24] V. P. Kotlyarov and E. Ya. Khruslov, Teor. Mat. Fiz. **68**, 172 (1986).

Translated by Eugene Yankovsky

Influence of quantum transitions in the continuum on ionization of atoms in strong fields

A. D. Kondorskiĭ and L. P. Presnyakov*)

P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 27 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1196–1209 (April 1999)

The tight-binding method is used to analyze the ionization of a hydrogenlike atom by an intense monochromatic laser field. The orthogonal and normalized basis in which the solution of the time-dependent Schrödinger equation is expanded contains unperturbed wave functions of the discrete spectrum and generalized Coulomb wave functions of the continuum. In the solution of the coupled equations we make use of the fact that the bound–free and free–free transitions are efficient in different regions of complex time. Simplified equations are constructed and investigated. Results of calculations for ionization of a hydrogen atom from its ground state and of the energy distribution of the electrons in strong and superstrong linearly polarized fields are presented. It is shown that in this case the ground state decays completely, and free–free transitions play a defining role in the dynamics of the process. It is established that the total probability of population of the upper Rydberg states abutting the continuum does not exceed 0.05. The range of applicability of the approach is discussed. A comparison with numerical results obtained by other authors is given.
[S1063-7761(99)00304-2]

1. INTRODUCTION

In an analysis of the ionization of atoms by an intense electromagnetic field or by ion impact, the influence of a time-dependent external potential (comparable to the intra-atomic potential or greater) must be taken into account even in the zeroth-order theory, i.e., in the wave functions of the continuum in the final reaction channel. The efficacy of such an approach follows from the ground-breaking work of Keldysh[1] and also from the subsequent development of the theory of ionization of atoms in a strong laser field.[2–8] A detailed analysis of a great deal of work on this subject is given by the review in Ref. 8, which shows under what physical conditions a theoretical description can be constructed with the help of various formulations of perturbation theory in combination with classical and semiclassical models. Thus, the Coulomb potential of the atomic core[9] in particular, which is not reflected in the basis wave functions, is taken into account.[1] In fields comparable to or exceeding the atomic fields, the application of perturbation theory is not well-founded, but a direct comparison of any calculation with the experimental data is hindered by the specifics of the experiment.[8]

In the physics of ion–atom collisions a comparison of theory with experiment is achievable with quantitative accuracy suitable for detailed checking of the theoretical results and methods. The intrinsic criteria of the theory developed in this way can also be used in the problem of the ionization of atoms in strong electromagnetic fields. We note above all that a generalization of the Keldysh theory[1] to the ionization of atoms by multiply-charged ions[10] which goes beyond perturbation theory turns out to be very promising. One result of this effort is a quantitative description of the experimental data on ionization of atoms[10] and neutralization of negative ions[11] by multiply-charged ions. Another important element of this approach is replacement of the generalized plane waves[1] by the Coulomb wave functions of the continuum, which take account of the Coulomb potential of the atomic core and the additional momentum acquired by the electrons an external, time-varying field.[12–15] These effects are important in single-electron[14] and two-electron ionization processes.[14,15] An analysis of the methods developed in the physics of atomic collisions shows that an important criterion for the applicability of the theory is that it be constructed in such a way that the S matrix is unitary at all stages of the calculations.

The present work addresses the influence of quantum transitions in the continuum on the ionization of atoms by a strong electromagnetic field. Free–free transitions have been separately studied by many authors in weak monochromatic fields on the basis of perturbation theory (see, e.g., Refs. 16–18). The influence of such transitions on ionization is taken into account to some degree in the results of numerical methods of solving the time-dependent Schrödinger equation for a negative ion[19] and a hydrogen atom,[20,21] as well as in numerical calculations based on the R-matrix theory of multiphoton processes.[22] A consistent theory should incorporate an analysis of direct and inverse processes. In this regard, the review in Ref. 23 is of interest, which considers radiation-stimulated recombination as the process inverse to multiphoton ionization. Perturbation-theory methods, mainly those discussed in Ref. 23, which take the specifics of such processes into account, nevertheless rather quite a complete description of the physical picture of the interrelated processes.

Below we utilize the method of tight binding of reaction

channels, which is based on an expansion of the solution of the time-dependent Schrödinger equation in orthonormal functions of the discrete and continuous spectra of the atom. Section 2 discusses the adopted basis and derives the tight-binding equations based on it. Section 3 shows that bound–free and free–free transitions are efficient on different intervals of the time axis (the t axis). On this basis a physical model is constructed which leads to simplified tight-binding equations. Methods for solving the simplified equations are discussed in Sec. 4. Results of calculations for ionization of a hydrogen atom from the ground state in strong and superstrong linearly polarized fields are presented in Sec. 5. The population of the higher Rydberg states is discussed, where Coulomb crowding of the discrete spectrum is treated as an analytic continuation of the continuum. It is shown that induced transitions within the continuum lead to total decay of the ground state, but the probability of population of the Rydberg states does not exceed 5%.

2. STATEMENT OF THE PROBLEM

We solve the problem of the ionization of a hydrogenlike atom in a classical monochromatic electromagnetic field with the help of the tight-binding method, i.e., by expanding the solution of the time-dependent Schrödinger equation over the full set of orthogonal and normalized basis functions of the discrete and continuous spectra. The generalized Coulomb waves[14] are orthogonal to the unperturbed functions of the discrete spectrum and have the form

$$\Psi_{\mathbf{k}}(\mathbf{r},t)=Q^{(-)}(\nu,\mathbf{p}(t),\mathbf{r})\exp\left\{-\frac{i}{2}\int_0^t p^2(\tau)\,d\tau\right\}. \quad (1)$$

Here $Q^{(-)}$ is the Coulomb function, containing in its asymptotic limit a converging wave and corresponding to motion of an electron in the field of a nucleus with charge Z and generalized momentum

$$\mathbf{p}(t)=\mathbf{k}-\mathbf{A}(t), \quad \nu=Z/p(t), \quad (2)$$

where $\mathbf{A}(t)$ is the vector potential and $\mathbf{k}$ is the ''unperturbed'' momentum. It is easy to see that for $Z=0$ the functions (1) coincide with the generalized Gordon–Volkov–Keldysh plane waves,[1] and for $\mathbf{A}(t)=0$ they go over to ordinary Coulomb wave functions of the continuum. The basis we use contains unperturbed wave functions in the discrete spectrum and generalized Coulomb waves (1) in the continuum.

The time-dependent Schrödinger equation in the velocity gauge

$$i\frac{\partial\Psi(\mathbf{r},t)}{\partial t}=\left[-\frac{1}{2}\Delta-\frac{Z}{r}+i\mathbf{A}(t)\cdot\nabla\right]\Psi(\mathbf{r},t), \quad (3)$$

should be solved with the standard initial condition

$$\Psi(\mathbf{r},t)=\varphi_0(\mathbf{r})\exp(-iE_0t_i), \quad (4)$$

where $\varphi_0(\mathbf{r})$ is the unperturbed wave function of the initial state with binding energy E_0. Representing the solution of Eq. (3) in the form

$$\Psi(\mathbf{r},t)=\sum_{s=0}^{\infty}a_s(t)\varphi_s(\mathbf{r})\exp(-iE_st)+\int a_{\mathbf{k}}(t)Q^{(-)}(\nu,\mathbf{p}(t),\mathbf{r})\,d\mathbf{k}, \quad (5)$$

where s is the set of quantum numbers of the discrete states, we obtain a Hermitian system of differential equations for the amplitudes a_s and $a_{\mathbf{k}}$ satisfying the normalization condition (unitarity of the S matrix)

$$\sum_{s=0}^{\infty}|a_s(t)|^2+\int|a_{\mathbf{k}}(t)|^2\,d\mathbf{k}=1. \quad (6)$$

Note that the original equation (3) can also be written in the length gauge, and Eqs. (4)–(6) retain the same form. The transformation from one representation to the other is realized simply by a phase factor. In an exact solution of the problem, the two representations lead to the same results. The use of approximations presupposes a choice of one of these gauges. To analyze free–free transitions, which play an important role in the present problem, the optimal choice is the velocity gauge (3).

Before discussing methods of solving the system of equations for the amplitudes, we first indicate the underlying approximations adopted in the present work. First, in the sum over discrete states on the right-hand side of Eq. (5), we retain only the ground state ($s=0$) and the Coulomb crowding (the upper Rydberg states) directly abutting the continuum, i.e., the sum from $s=s_m\gg1$ to $s\to\infty$. Second, we neglect transitions from the ground state to these highly excited bound states and consider their coupling only with the continuum. The first approximation leads to the neglect of quantum transitions by way of intermediate levels, and is often used in various approaches.[3–8] The second corresponds to the preferential population of the higher Rydberg states via the continuum.[24,25] We consider these states to be an analytic continuation of the continuum, and we do not write the sum over the Coulomb crowding separately in the formulas of this section, denoting it together with the integral over the continuum by a single integral sign.

The Hermitian system of equations for the amplitudes that follows from substituting the functional dependence (5) in Eq. (3) takes the form

$$\begin{aligned}i\dot a_0(t)&=\mathbf{A}(t)\int a_{\mathbf{k}}(t)\mathbf{U}_{0,\mathbf{k}}(t)\,d\mathbf{k},\\ i\dot a_{\mathbf{k}}(t)&=\mathbf{A}(t)\left\{\mathbf{U}^*_{0,\mathbf{k}}(t)a_0(t)+\int\mathbf{U}_{\mathbf{k},\mathbf{k}'}(t)a_{\mathbf{k}'}(t)\,d\mathbf{k}'\right\}.\end{aligned} \quad (7)$$

The matrix elements are

$$\mathbf{U}_{0,\mathbf{k}}(t)=\exp\left\{-\frac{i}{2}\int_0^t(p^2(\tau)+2|E_0|)\,d\tau\right\}\times\int\varphi_0(\mathbf{r})\nabla Q^{(-)}(\nu,\mathbf{p}(t),\mathbf{r})\,d\mathbf{r}, \quad (8)$$

$$\mathbf{U}_{\mathbf{k},\mathbf{k}'}(t)=\exp\left\{-\frac{i}{2}\int_0^t (p'^2(\tau)-p^2(\tau))\,d\tau\right\} \times\int Q^{(-)*}(\nu,\mathbf{p}(t),\mathbf{r})\,\nabla Q^{(-)}(\nu',\mathbf{p}'(t),\mathbf{r})\,d\mathbf{r}, \tag{9}$$

$$\mathbf{U}_{\mathbf{k},\mathbf{k}'}(t)=\mathbf{U}^*_{\mathbf{k}',\mathbf{k}}(t). \tag{10}$$

The integrals in expressions (8) and (9), namely the integral of the bound–free (8) and free–free (9) dipole transitions, can be expressed in analytic form.[26] The initial condition (4) reads

$$a_0(t_i)=1, \quad a_{\mathbf{k}}(t_i)=0. \tag{11}$$

In what follows we restrict attention to linearly polarized monochromatic fields with constant field strength **F**:

$$\mathbf{A}(t)=-\frac{\mathbf{F}}{\omega}\sin\omega t. \tag{12}$$

3. PHYSICAL MODEL AND SIMPLIFIED EQUATIONS

We analyze the properties of the solutions of system (7) in the generalized momentum $[\mathbf{p}(t)]$ representation (2). The selection rules for bound–free transitions (8) are the same as in the original theory:[1] the ground state of a hydrogenlike atom with orbital angular momentum $l=0$ and projection $m=0$ interacts with all states of the continuum with $l=1$ and $m=0$. The latter rule is a rigorous consequence of the linear polarization of the field (12). Transitions to states of the continuum with $l\neq 1$ take place only as a consequence of an interaction within the continuum itself. Transitions to states of the continuum with $l=1$ from the ground state take place for complex values of the time t_0^j defined by the condition that the derivative of the phase in expression (8) vanish,

$$p^2(t_0^j)+2|E_0|=0, \quad j=0,1,2,\ldots,N, \tag{13}$$

where the value of t_0^j depends on the value of the ''unperturbed'' momentum $\mathbf{k}$.[1,7,10] In the terminology of the quantum theory of nonadiabatic transitions, this corresponds to promotion of the term (energy level) of the ground state to the continuum. States of the continuum with $l=1$ are populated efficiently in the immediate vicinity of the points $t_0^j(\mathbf{k})$. Between the points t_0^j and t_0^{j+1}, separated on the real t axis by a half-period of oscillation of the monochromatic field (12), the probabilities of bound–free transitions are small in comparison with the probability of transitions within the continuum. The above arguments lead to a model similar to one widely used in the physics of atomic collisions.[27]

We partition the real t axis into intervals of length $\mathrm{Re}\,t_0^j-\Delta t<t<\mathrm{Re}\,t_0^j+\Delta t$, corresponding to maximum efficiency of the bound–free transitions, and intervals between these regions, $\mathrm{Re}\,t_0^j+\Delta t<t<\mathrm{Re}\,t_0^{j+1}-\Delta t$, corresponding to efficient transitions within the continuum (free–free transitions). In the regions of maximum efficiency of the bound–free transitions we solve the system of equations

$$i\dot a_{\mathbf{k}}=\mathbf{A}(t)\mathbf{U}^*_{0,\mathbf{k}}(t)a_0,$$
$$i\dot a_0=\mathbf{A}(t)\int \mathbf{U}_{0,\mathbf{k}}a_{\mathbf{k}}d\mathbf{k}. \tag{14}$$

In the regions of maximum efficiency of the free–free transitions the system of equations has the form

$$i\dot a_{\mathbf{k}}=\mathbf{A}(t)\int \mathbf{U}_{\mathbf{k},\mathbf{k}'}a_{\mathbf{k}'}d\mathbf{k}'. \tag{15}$$

The difference from the *S*-matrix joining method[27] is that Eq. (15) leads to a change not only in the phases but also in the absolute values of the amplitudes on the intervals between the regions of bound–free transitions .

We call system (14) and (15) the system of simplified equations. The approximations employed to obtain it are consistent with unitarity of the *S* matrix:

$$|a_0(t)|^2+\int |a_{\mathbf{k}}(t)|^2\,d\mathbf{k}=1. \tag{16}$$

The initial values of a_0 and $a_{\mathbf{k}}$ on each interval follow in a natural way from their final values in the preceding interval.

4. SOLUTION OF THE SIMPLIFIED EQUATIONS

Solution of the simplified equations divides into two parts. First of all, we must solve problem (14) in order to describe the behavior of the system in the region of the nonadiabatic transition. It is convenient to carry out an approximate integration over the continuum already in system of equations (14) thereby reducing it to two ordinary differential equations.

We make the substitution

$$a_{\mathbf{k}}(t)=c_{\mathbf{k}}(t)b(t), \quad \int |a_{\mathbf{k}}(t)|^2 d\mathbf{k}=|b(t)|^2,$$
$$\int |c_{\mathbf{k}}(t)|^2 d\mathbf{k}=1. \tag{17}$$

Let us determine the form of the coefficients of the desired system whose solution optimally approximates the value of a_0 following from system (14). Writing

$$c_{\mathbf{k}}(t)=s_{\mathbf{k}}(t)\exp\left\{i\int^t \alpha_{\mathbf{k}}(\tau)\,d\tau+2i\int^t \beta(\tau)\,d\tau\right\}, \tag{18}$$

the desired system becomes

$$i\dot a_0=-\beta a_0+Wb,$$
$$i\dot b=\beta b+W^* a_0, \tag{19}$$

where

$$W(t)=\left(\int U_{\mathbf{k},0}(t)c^*_{\mathbf{k}}d\mathbf{k}\right) \times\exp\left\{i\int^t d\tau\left[\int \alpha_{\mathbf{k}}(\tau)s^2_{\mathbf{k}}(\tau)d\mathbf{k}\right]\right\}, \tag{20}$$

and the coefficient β can be calculated by solving the quadratic equation

$$\int |U_{0,\mathbf{k}}(t)|^2 d\mathbf{k} = |W|^2 + \left(\frac{d}{dt}\arg W^*\right)\beta + \beta^2, \tag{21}$$

$$\beta = -\frac{\Delta}{2} + \sqrt{\frac{\Delta^2}{4} + \int |U_{0,\mathbf{k}}|^2 d\mathbf{k} - |W|^2}, \tag{22}$$

$$\int_0^t \Delta(\tau)\, d\tau = \arg W^*. \tag{23}$$

Analysis shows that for $c_{\mathbf{k}}$ we must use the normalized contribution of the singular point of the matrix element to the integral of this matrix element over time. This gives the following expressions for $s_{\mathbf{k}}(t)$ and $\alpha_{\mathbf{k}}(t)$:

$$s_{\mathbf{k}}(t) = \sqrt{\frac{F}{2\sqrt{(2|E_0|)^3}}}\exp\left\{-\frac{\mathbf{k}_\perp\sqrt{2|E_0|}}{2F}\right\}\delta(\mathbf{k}_\parallel), \tag{24}$$

$$\alpha_{\mathbf{k}}(t) = \left(\frac{\mathbf{k}^2}{2} - E_0 + \frac{F^2}{8\omega^2|E_0|}\right) t. \tag{25}$$

The coefficient β follows from formula (22).

Note that the integral

$$\int |U_{0,\mathbf{k}}(t)|^2\, d\mathbf{k} = \int |U_{0,\mathbf{p}}(t)|^2\, d\mathbf{p} \tag{26}$$

is proportional to the sum of oscillator strengths over the entire continuum. When the coefficient $|W|^2$ in formula (22) is close to or equal to the quantity in (26), $\beta = 0$ and system (19) coincides with the system introduced phenomenologically in Ref. 28.

Analysis shows that even in the simple case of a linearly polarized field (12), the behavior of the solution of the initial system (14) and of the solution of the approximate system (19) in the vicinity of the singular points leads to the impossibility of using the well-known methods of reference equations and other analytic models. In what follows, we investigate system (19) numerically.

The second step in the solution an analysis of the equations for the interactions within the continuum (15). The system of tight-binding equations, upon expansion over eigenvalues of the angular momentum l (in the $\mathbf{p}$ representation), has the form

$$i\dot{a}_{l,\mathbf{k}} = \sum_{l'} \int U_{l,\mathbf{k},l',\mathbf{k}'} a_{l',\mathbf{k}'} d\mathbf{k}. \tag{27}$$

Since the wave functions in our basis depend on $\mathbf{p} = \mathbf{k} - \mathbf{A}$, it is convenient to transform from the variables $\mathbf{k}$ to the variables $\mathbf{p}$. Since $|p^2 - p'^2| \ll p^2$ under the assumptions of the model considered here, the matrix elements of the free–free transitions (the so-called Sommerfeld–Nordsieck integrals) can be approximately expressed in terms of Hankel functions.[29] In this case we have made direct use of the Sommerfeld method, based on the integral representation of the hypergeometric functions. It is well known[26] that at large values of the orbital angular momentum the Sommerfeld method[29] yields the semiclassical results. However, in our work, small and intermediate values of l, for which the semiclassical approach is not justified, play an important role. We make the subsitution of variables

$$a_{l,\mathbf{p}} = \exp\left\{-i\mathbf{k}\int_0^t \mathbf{A}(\tau)\, d\tau\right\}\tilde{a}_{l,\mathbf{p}}. \tag{28}$$

If we now transform in the independent variable to the energy $E = p^2/2$ and invoke the property $\mathbf{p} \approx \mathbf{A}$, then we can rewrite the system of tight-binding equations in the form

$$i\dot{a}_{l,E} = \sum_{l'} \int_{-E_n}^{\infty} W_{l,l'}(E - E', t) a_{l',E'} dE',$$

$$W_{l,l'}(E - E', t) = W^*_{l,l'}(E' - E, t), \tag{29}$$

$$W_{l,l\pm1}(E - E', t) = \frac{32\pi^2\omega^2|E_0|}{F^2\sin^2\omega t} \times \sqrt{\left(l + 1 \pm \frac{1}{2}\right)(2l+1)(2l+1\pm2)} \times \exp[i(E - E')t] \times \left(\frac{\partial}{\partial x_\pm} H^{(1)}_{i\nu}(x_\pm) - i\,\mathrm{sign} \times (E - E') H^{(1)}_{i\nu}(x_\pm)\right), \tag{30}$$

$$\nu = \frac{|E - E'|\omega^3\sqrt{2|E_0|}}{F^3\sin^3\omega t},$$

$$x_\pm = i\frac{|E - E'|\omega^2}{F^2\sin^2\omega t}\sqrt{(l \pm 1/2)(l + 1 \pm 1/2)}. \tag{31}$$

As noted earlier, we are considering highly excited states of the atom to be a continuation of the continuum to negative energies. By virtue of the rapid convergence of the integral (29) at its lower limit, the value of E_n can tend to $-\infty$. The rapid convergence of the integral is related to population of the Coulomb crowding (the upper Rydberg states) from the continuum and to the abrupt falloff in the population of these states with decreasing principal quantum number.

The equations written in this form can be solved by quadratures. We define the column vector $(a_{0,E}, a_{1,E}, \ldots)^T = \overline{a_E}$ and the matrix $W_{l,l'}(E - E', t) = \hat{W}(E - E', t)$. Then, according to the rules of matrix multiplication,

$$i\dot{\overline{a_E}} = \int_{-\infty}^{\infty} \hat{W}(E - E', t)\, \overline{a_{E'}}\, dE'. \tag{32}$$

We introduce the generating function in the form of a column vector:

$$\overline{G}(\varphi, t) = \int_{-\infty}^{\infty} \overline{a_\xi}(t) e^{-i\varphi t} d\varphi. \tag{33}$$

It satisfies the equation

$$i\dot{\overline{G}} = \left(\int_{-\infty}^{\infty} e^{-i\varphi x}\hat{W}(x, t) dx\right)\overline{G}, \tag{34}$$

which has a solution in the form of a matrix exponential

$$\overline{G}(\varphi,t)=\exp\left(-i\int_0^t d\tau\int_{-\infty}^{\infty}e^{-i\varphi x}\hat{W}(x,\tau)dx\right)\overline{G}(\varphi,0). \tag{35}$$

The column-vector solution follows from the generating function by the inverse Fourier transformation

$$\overline{a}_E(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\overline{G}(\varphi,t)e^{i\varphi t}d\varphi. \tag{36}$$

As the region of nonadiabaticity in the vicinity of the point $t_j=\pi j$ in the continuum is traversed, the coefficients $a_{l,E}$ undergo the following changes:

$$a_{l,E}(t_j+0)=a_{l,E}(t_j-0)+\delta_{l,1}\delta b_j\sigma_j(E), \tag{37}$$

where $\delta b_j=b(t_j+0)-b(t_j-0)$. Here $b(t)$ is the corresponding coefficient of system (19) and $\sigma_j(E)$ is the normalized distribution with which the transition to (and from) the ground state takes place. The initial conditions for system (19) at each point $t_j=\pi j$ have the form

$$a(t_j-0)=a(t_{j-1}+0)e^{i\eta_j}, \tag{38}$$

$$b(t_j-0)=\int_0^{\infty}b_{1,E}(t_j-0)\sigma_j^*(E)dE, \tag{39}$$

where η_j is the adiabatic phase shift of the coefficient a on the segment t_{j-1},t_j, and

$$\sigma_j(E)=\sqrt{\frac{F}{2\sqrt{(2|E_0|)^3}}}\exp\left\{-\frac{E\sqrt{2|E_0|}}{F}+i\frac{\pi}{\omega}\left(E-E_0+\frac{F^2}{8\omega^2|E_0|}\right)\right\}. \tag{40}$$

Results of numerical calculations following this scheme are given below. The results can be represented in another form. Consider the column vectors $\overline{a}_j=(a(t_j-0),b(t_j-0))^T$. Then the column vectors corresponding to different j are related by the equation

$$\overline{a}_j=\hat{N}_j\hat{S}\overline{a}_{j-1}, \tag{41}$$

where $\hat{S}$ is the matrix of the nonadiabatic transition and $\hat{N}_j$ is a matrix describing the variation of $\overline{a}(t)$ on the interval (t_{j-1},t_j). If we assume that $\hat{N}_j$ depends weakly on j (as confirmed by calculation), then by writing the matrices in the form

$$\hat{S}=\begin{Vmatrix}\sqrt{1-\Pi^2}\,e^{iz} & i\Pi\\ i\Pi & \sqrt{1-\Pi^2}\,e^{-iz}\end{Vmatrix},$$

$$\hat{N}=e^{i\chi}\begin{Vmatrix}1 & 0\\ 0 & e^{-\mu+ix}\end{Vmatrix}, \tag{42}$$

where Π, z, μ, x, and χ are real quantities, it is possible to obtain a common approximate expression for a_j:

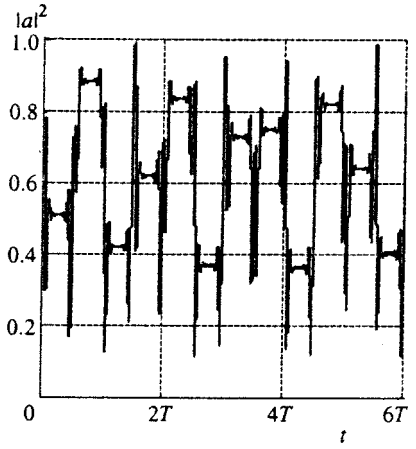

FIG. 1. Numerical solution of (19) for $F/\omega\gg 1$ ($F=0.3$ a.u., $\omega=0.1$ a.u.). The figure shows the probability of finding an electron in the ground state as a function of time (T is the period of the external field).

$$a_j=\frac{\exp[(-\mu+ix)j/2]}{2g}(\sqrt{1-\Pi^2})^j\cos^{j-1}\varphi\times\{(1+g)^j(i\sin\varphi+g\cos\varphi)-(1-g)^j\times(i\sin\varphi-g\cos\varphi)\},$$

$$g=\sqrt{1-\frac{1}{(1-\Pi^2)\cos^2\varphi}},\quad \varphi=z-\frac{x+i\mu}{2}. \tag{43}$$

The presence of the quantity μ in this formula and the fact that the quantity φ is complex reflect the influence of free–free transitions on states with high energies and $l\neq 1$.

5. RESULTS AND DISCUSSION

In the present section we present some results of a numerical study of the solution of the simplified equations described above. Above all, note that the solution of the simplified Hermitian system without allowance for interactions within the continuum (19) does not describe total ionization of an atom for arbitrary values of the parameters (time, field strength, frequency of the field). Free–free transitions, described by Eqs. (27), lead to population of states with $l\neq 1$ and, accordingly, to a drain of probability from states that interact indirectly with the ground state. It is specifically as a result of quantum transitions within the continuum that total ionization of the atom is possible.

Figure 1 plots the time dependence of the population of the ground state of a hydrogen atom, calculated taking into account free–free transitions with $F/\omega\gg 1$. As can be seen from the figure, interaction between levels occurs in a small neighborhood of the points $t_j=\pi j$, $j=0,1,2,\ldots,N$, which confirms the applicability of the present approximation in this case. Thus, in the given case the ground state does not decay. If interactions within the continuum are taken into account, the picture changes. Figure 2 plots the time dependence of the population of the ground state of the same atom taking into account transitions within the continuum.

It can be seen by comparing the figures that free–free transitions lead to total decay of the ground state. A charac-

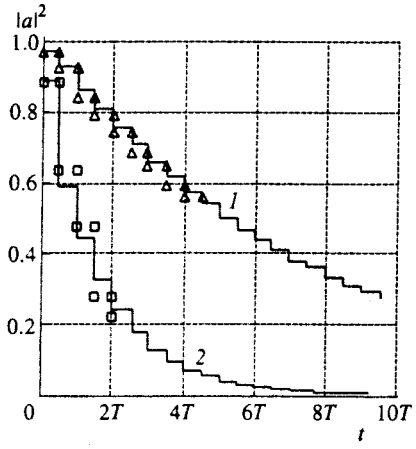

FIG. 2. Population of the ground state as a function of time (T is the period of the external field) for $\omega=0.1$ a.u.: *1* — $F=0.2$ a.u., *2* — $F=0.4$ a.u. The points were calculated using Eqs. (17)–(40), the solid curves represent the approximat ion (43) with the matrix $\hat{N}$ averaged over several initial points.

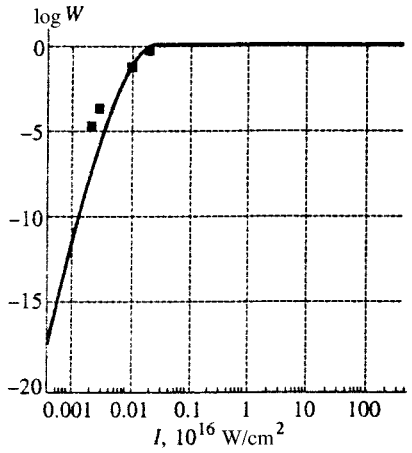

FIG. 4. Dependence of the probability of ionization during 25 fs on the intensity of the external field for $\omega=2$ eV. Filled squares — as calculated in Ref. 21, solid line — calculated in the present work.

teristic graph of the distribution over l is shown in Fig. 3. In addition, the probability that an electron will return to the ground state can decrease if its energy is too high. This effect can be characterized by the ratio of the population of the upper level of a two-level system to the total probability that the electron will be found in states of the continuum with $l=1$. The value of this parameter turns out to depend weakly on the field, and is roughly 75%.

Figure 4 plots the dependence we obtained (solid line) of the ionization probability of the ground state of a hydrogen atom during 25 fs on the intensity of the external field for $\omega=2$ eV, along with the result (squares) of direct numerical calculation.[21] The given approach allows one to find the time dependence of the population of the states of an atom for fields comparable to or exceeding the atomic field. The results correspond to parameter values corresponding to tunneling ionization and above-threshold ionization. The discrepancy in the results at low intensities ($F/\omega\ll 1$) results from dividing the time intervals, in deriving the simplified equations (14), (15), into intervals of maximum efficiency for bound–free transitions and free–free transitions. Analysis shows that the given division is applicable for values of the parameter $F/\omega\geqslant 1$. Although system (19) is also applicable for $F/\omega\ll 1$, a study of weak fields is not the goal of the present work.

Calculation for large values of the external field shows that at intensities $I>3\times 10^{14}$ W/cm^2 ($F>0.1$ a.u.), essentially total ionization occurs by the end of the second period of the external field. Figure 5 plots the dependence of the total population of the continuum after the first half-period of the external field on the intensity of the external field. As can be seen, the population oscillates about a constant value. Knowing the properties of the solution of the equations for free–free transitions (the matrix $\hat{N}$) and employing Eq. (43), it is possible to describe the ionization probability per unit time for external fields many times stronger than the atomic field. The latter also oscillates about a constant value as the field intensity increases.

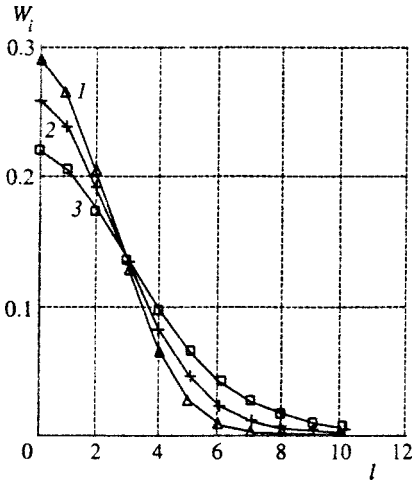

FIG. 3. Distributions over l in the continuum at various times: *1* — $t=T/2$, *2* — $t=T$, *3* — $t=3T/2$. The case $\omega=0.1$ a.u., $F=0.2$ a.u.

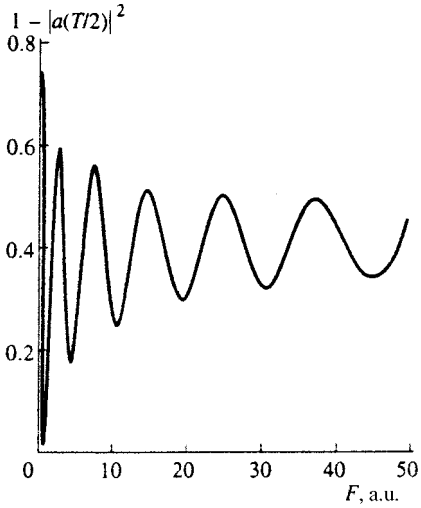

FIG. 5. Dependence of the total population of the continuum after the first half-period of the external field as a function of external field strength ($\omega=0.1$ a.u.).

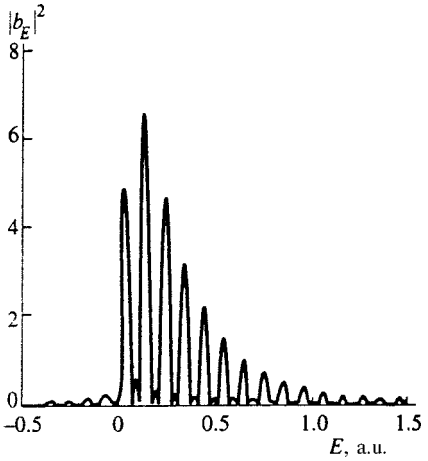

FIG. 6. Energy distribution in the continuum ($t=3T/2$, $F=0.5$ a.u., $\omega=0.1$ a.u.).

In addition, we found the energy distributions of the electrons in the continuum and in the Coulomb crowding at various times. Figure 6 displays a typical distribution for a hydrogen atom, consisting of a set of equidistantly spaced peaks, the distance between which is equal to ω. This aspect of the curve is a result of taking a large number of points t_0^j into account (see Sec. 3), each of which introduces its own phase shift (40). The above picture agrees with the results of other authors.[21]

From Fig. 6 it can be seen that the population of the upper Rydberg states falls off abruptly with distance from the edge of the continuum, i.e., with decreasing principal quantum number. This result confirms the realism of the assumptions used in the solution of the simplified equations (29), (32).

The ratio of the total population of the Rydberg levels of the hydrogen atom to the population of the entire continuum is plotted in Fig. 7 as a function of electric field strength.

6. CONCLUSION

An important element of the above analysis is the use of a basis of generalized Coulomb functions of the continuum to solve the time-dependent Schrödinger equation by the tight-binding method. Wave functions similar in their meaning to the unperturbed wave functions of the discrete spectrum (1) are also proposed in Refs. 12 and 13; however, they are not orthogonal to the wave functions of the discrete spectrum, which hinders their use in the tight-binding method. Indeed, a set of non-orthogonal basis functions complicates the resulting tight-binding equations due to the appearance of the non-orthogonality matrix on the left-hand side of the equations.

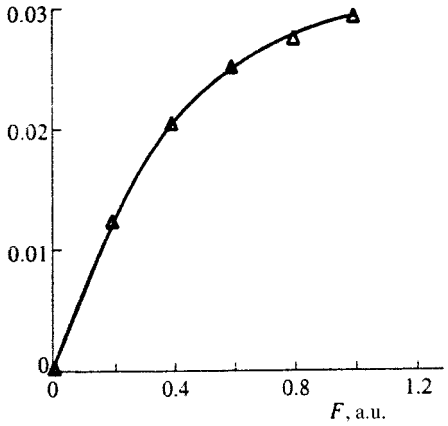

FIG. 7. Average ratio of the total population of the Rydberg states to the total population of the continuum versus external field strength ($\omega=0.1$ a.u.).

Using this method it is easy to see that quantum transitions between states of the continuum play a decisive role in the ionization of atoms in a strong field. The principal contribution to this process comes from transitions between continuum states with orbital angular momentum $l=1$ and other states of the continuum. Neglecting such transitions leads to oscillations in the population of the ground state of the atom as a function of the physical parameters of the problem, even in superstrong fields. Strictly speaking, the atom does not ionize completely, since levels of the Coulomb crowding (higher Rydberg states) directly abutting the continuum turn out to be populated. However, their total probability of occupancy is at most 5%.

We are grateful to N. B. Delone, S. P. Goreslavskiĭ, D. F. Zaretskiĭ, V. P. Kraĭnov, and M. V. Fedorov for fruitful discussions. The main results of this work were discussed in detail with D. B. Uskov, to whom we are deeply grateful for a number of valuable comments. We are grateful to the reviewer for a constructive analysis and valuable remarks. This work was partially supported by the Russian Fund for Fundamental Research (Grant No. 96-16090) and INTAS (Project No. 94-4698).

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

[1] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1965)].
[2] A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **50**, 1393 (1966) [Sov. Phys. JETP **23**, 924 (1966)].
[3] F. H. M. Faisal, J. Phys. B **6**, L89 (1973).
[4] H. R. Reiss, Phys. Rev. A **22**, 1786 (1980); **42**, 1476 (1990).
[5] Yu. A. Bychkov and A. M. Dykhne, Zh. Éksp. Teor. Fiz. **58**, 1734 (1970) [Sov. Phys. JETP **31**, 928 (1970)].
[6] D. F. Zaretskiĭ and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz. **66**, 537 (1974) [Sov. Phys. JETP **39**, 257 (1974)]; ibid. **67**, 1301 (1975) [Sov. Phys. JETP **40**, 647 (1975)].
[7] N. B. Delone and V. P. Kraĭnov, *Atoms in Strong Light Fields* [in Russian], Atomizdat, Moscow (1978) [Springer Series in Chemical Physics, Vol. 28, Springer-Verlag, Berlin (1985)].
[8] N. B. Delone and V. P. Kraĭnov, Usp. Fiz. Nauk **168**, 531 (1998).
[9] V. P. Krainov, J. Opt. Soc. Am. B **14**, 425 (1997).
[10] L. P. Presnyakov and D. B. Uskov, Zh. Éksp. Teor. Fiz. **86**, 882 (1984) [Sov. Phys. JETP **59**, 515 (1984)].
[11] F. Melchert, M. Brenner, S. Krüdener, R. Schulze, S. Meuser, K. Huber, E. Salzborn, D. B. Uskov, and L. P. Presnyakov, Phys. Rev. Lett. **74**, 888 (1995).
[12] J. Z. Kamiński, Physica Scripta A **42**, 417 (1990).
[13] M. H. Mittleman, Phys. Rev. A **50**, 3249 (1994).
[14] L. P. Presnyakov, H. Tawara, I. Yu. Tolstikhina, and D. B. Ustov, J. Phys. B **28**, 785 (1995).
[15] L. P. Presnyakov and D. B. Uskov, JETP Lett. **66**, 22 (1997).
[16] B. A. Zon, Zh. Éksp. Teor. Fiz. **73**, 128 (1977) [Sov. Phys. JETP **46**, 65 (1977)].
[17] M. H. Mittleman, Phys. Rev. A **14**, 1338 (1976); **16**, 1549 (1977); **21**, 79 (1980).

[18] F. W. Byron, Jr., P. Francken, and C. J. Joachain, J. Phys. B **20**, 5487 (1987).
[19] L. A. Collins and A. L. Merts, Phys. Rev. A **45**, 6615 (1992).
[20] J. Zhang and P. Lambropoulos, J. Nonlinear Opt. Phys. Mater. **4**, 633 (1995).
[21] E. Cormier and P. Lambropoulos, J. Phys. B **30**, 77 (1997).
[22] C. J. Joachain, M. Dorr, and N. J. Kylstra, Comments At. Mol. Phys. **33**, No. 5, 247 (1997).
[23] F. Ehlotsky, A. Jaroń, and J. Z. Kamiński, Phys. Rep. **297**, 63 (1998).
[24] M. V. Federov and A. M. Movsesian, J. Opt. Soc. Am. B **6**, 928 (1989).
[25] A. M. Movsesyan and M. V. Federov, Zh. Éksp. Teor. Fiz. **95**, 42 (1989) [Sov. Phys. JETP **68**, 27 (1989)].
[26] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P, Pitaevskiĭ, *Quantum Electrodynamics*, 2nd ed. Pergamon Press, Oxford (1982).
[27] E. E. Nikitin and S. Ya. Umanskiĭ, *Non-Adiabatic Transitions During Slow Atomic Collisions* [in Russian], Atomizdat, Moscow (1979).
[28] R. K. Janev and L. P. Presnyakov, J. Phys. B **13**, 4233 (1980).
[29] A. Sommerfeld, *Atomic Structure of Spectral Lines*, 3rd ed., transl. from the 5th German ed., Methuen, London (1934).

Translated by Paul F. Schippnick

Sub-Poissonian radiation of a one-atom two-level laser with incoherent pumping

A. V. Kozlovskiĭ*) and A. N. Oraevskiĭ

P. N. Lebedev Physics Institute, 117924 Moscow, Russia
(Submitted 25 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1210–1220 (April 1999)

The quantum statistical properties of the radiation of a one-atom two-level laser with incoherent pumping are analyzed. Solution of the Liouville equation for the density operator in the basis of Fock states shows that stationary radiation from a single-mode laser with incoherent pumping can be in a squeezed (sub-Poissonian) stationary state if the rate of spontaneous decay is lower than the rate of cavity losses and the pump rate. Inside the cavity the Fano factor reaches $F=0.85$ (15% squeezing). Multiple squeezing ($F=0.19$) is possible in the transient lasing regime. Significant squeezing obtains at the cavity output; the spectral Fano factor at zero frequency is 0.36 under optimal conditions. © *1999 American Institute of Physics.* [S1063-7761(99)00404-7]

1. INTRODUCTION

Microscopic laser systems consisting of an ion or atom held in a trap inside a high-Q resonator are presently the object of intense study. Recent detailed analyses[1–7] show that radiation in a resonant mode of such an elementary laser system exhibits a number of nonclassical properties such as sub-Poissonian statistics,[1–4], photon antibunching,[5] multipeaked Mollow fluorescence spectra for incoherent pumping,[6] thresholdless lasing,[3,5] and the corresponding photon antibunching and broadening of the spectral line in the many-level case.[5]

In the present paper we investigate one particular aspect of a one-atom laser, namely the sub-Poissonian statistics of the photon number of the radiation in the single-mode case for incoherent pumping. Recently, several techniques have been found for creating states of the electromagnetic field with fluctuations of the photon number below the shot-noise limit (sub-Poissonian light). Toward this end, Refs. 8–10 suggest the use of a single-mode laser with regularized pumping. References 11 and 12 use atomic coherence of the active medium to obtain squeezed light. Multilevel schemes of an ordinary single-mode laser, as shown in Refs. 13–19, can also be a source of squeezed light in the case of incoherent as well as coherent pumping. References 20 and 21 point out the possibility of generating sub-Poissonian light with the aid of a multimode laser. The conditions for creating a transient squeezed state of the generated radiation were found in Ref. 22 for a single-mode two-level laser with incoherent pumping. In the single-mode two-level lasers considered in Refs. 2, 5–7, 19, 22, and 23, considerable squeezing under steady-state conditions is absent for one-atom lasers as well as for multiatom lasers ($N_{\text{atom}}\gg 1$).[5–7]

In the present paper we analyze the dynamics of the quantum statistical properties of laser radiation with the help of the reduced density operator of the atom + field system in the basis of Fock states of the field. This analysis is valid for all possible relative values of the parameters characterizing the processes of incoherent pumping and dissipation, modeled with the help of the corresponding reservoirs interacting with the atom and the radiation field. Note that our studies of laser dynamics and the statistical properties of laser radiation do not make any small-fluctuation assumptions, in which it is customary to adopt a linear approximation in the fluctuations.

2. LASER MODEL

The Hamiltonian of the atom + field + reservoir system includes energy operators of the electromagnetic field (H_F), the atom (H_A), the interaction of the field with the atom (V_{A-F}), the energy of the reservoir of the continuous spectrum of thermal modes of the electromagneti c field (thermostat) (R_F), the interaction operator of the atom with the thermostat reservoir (V_{A-R}), the interaction operator of the cavity field with the thermostat (V_{F-R}), and the energy of the pump reservoir (R_P) and its interaction with the atom (V_{A-P}):

$$\frac{H}{\hbar}=(H_F+H_A+V_{A-F}+R_F+V_{A-R}+V_{F-R}+R_P+V_{A-P})\frac{1}{\hbar}$$

$$=\omega_c a^+a+\frac{\omega_A}{2}\sigma^z+g(a^+\sigma^-+\sigma^+a)+\sum_{j=1}^{\infty}\omega_j b_j^+ b_j$$

$$+\sum_{j=1}^{\infty} g_j(b_j^+\sigma^-+b_j\sigma^+)+\sum_{j=1}^{\infty} k_j(a^+b_j+b_j^+a)$$

$$+\sum_k \omega_k\Pi_k^+\Pi_k+\sum_k \mu_k(\Pi_k^+\sigma^-+\sigma^+\Pi_k). \qquad (1)$$

Here a (a^+) are the annihilation (creation) operators of the electromagnetic field of a discrete cavity mode with frequency ω_c; the operators b_j (b_j^+) are the annihilation (creation) operators of the reservoir of the continuous spectrum of thermal modes of the electromagnetic field present inside the cavity as a consequence of partial transparency of the mirrors. The indicated operators obey the commutation relations for Bose particles:

$$[a,a^+]=1, \quad [a^+,a^+]=[a,a]=0,$$

$$[b_j,b_k^+]=\delta_{jk}, \quad [b_j^+,b_k^+]=[b_j,b_k]=0. \tag{2}$$

The operators Π_k (Π_k^+) correspond to the field of the pump reservoir, which in general is not in thermodynamic equilibrium and consists of particles obeying Fermi statistics (e.g., electrons of the discrete or continuous energy spectrum), and obey the commutation relations

$$\{\Pi_k,\Pi_l^+\}=\delta_{kl}, \quad \{\Pi_k^+,\Pi_l^+\}=\{\Pi_k,\Pi_l\}=0, \tag{3}$$

where $\{\ldots\}$ is the anticommutator.

The term $\{b_j\}$ in Eq. (1), which represents the interaction of the laser radiation field with the reservoir modes. is responsible for setting up thermodynamic equilibrium of the cavity mode with the reservoir–thermostat field entering through the mirrors and having mean photon number in the mode[24,25]

$$n_T\equiv\overline{n}(\omega=\omega_c,T)=[\exp(\hbar\omega_c/kT)-1]^{-1}. \tag{4}$$

The atomic operators of polarization and population inversion of a two-level atom ($|\downarrow\rangle$ and $|\uparrow\rangle$ are respectively the upper and lower energy states of the active electron)

$$\sigma^-=|\downarrow\rangle\langle\uparrow|, \quad \sigma^+=|\uparrow\rangle\langle\downarrow|, \quad \sigma^z=|\uparrow\rangle\langle\uparrow|-|\downarrow\rangle\langle\downarrow| \tag{5}$$

satisfy the relations

$$2\sigma^{\mp}\sigma^{\pm}=(1\mp\sigma^z), \quad \sigma^z\sigma^z=1, \tag{6}$$

and the commutation relations for Fermi particles:

$$\{\sigma^+,\sigma^-\}=|\uparrow\rangle\langle\uparrow|+|\downarrow\rangle\langle\downarrow|=1, \quad \{\sigma^+,\sigma^+\}=\{\sigma^-,\sigma^-\}=0. \tag{7}$$

The term representing the interaction of the atomic electron with the field of harmonic oscillators of the thermostat $\{b_j\}$ ensures the establishment of thermal equilibrium of the quantum-averaged atomic operators with temperature T. The interaction term of the atomic polarization operators with the pump field $\{\Pi_k\}$ leads the system to deviate from equilibrium.

The interaction of the radiation field with the polarization of the atom is proportional to the coupling constant g determined by the transition dipole moment $d_{\uparrow\downarrow}$ and the volume of the cavity V

$$g=d_{\uparrow\downarrow}\sqrt{\frac{2\pi\omega_A}{\hbar V}}. \tag{8}$$

Similarly we define the coupling constants g_j of the electromagnetic field of the thermostat with an atomic electron. The parameters k_j and μ_k are respectively the coupling constants of the laser mode of the cavity field with the thermostat and the coupling constants of the atom with the pump reservoir.

We analyze the quantum stochastic dynamics of a two-level one-atom laser with the help of the reduced density operator of the atom + single-mode field system in the basis of Fock states

$$\rho(t)=\sum_{i,j=\{\uparrow,\downarrow\}}\sum_{n,m=0}^{\infty}\rho_{n,i;m,j}(t)|i\rangle|n\rangle\langle m|\langle j|. \tag{9}$$

In the interaction representation and in the Born–Markov approximation,[24,25] the reduced density operator of the atom + field system (9) interacting with a reservoir satisfies the Liouville equation

$$\begin{aligned}\frac{\partial\rho}{\partial t}=\mathrm{Tr}_R(\sigma_{R\oplus S})=&-i\Delta\left[\frac{\sigma^z}{2},\rho\right]-ig[(a^+\sigma^-+\sigma^+a),\rho]\\&+\frac{\gamma}{2}(n_T+1)(2a\rho a^+-a^+a\rho-\rho a^+a)\\&+\frac{\gamma}{2}n_T(2a^+\rho a-aa^+\rho-\rho aa^+)+\frac{\Gamma}{2}(N_T+1)\\&\times(2\sigma^-\rho\sigma^+-\sigma^+\sigma^-\rho-\rho\sigma^+\sigma^-)\\&+\frac{\Gamma}{2}N_T(2\sigma^+\rho\sigma^--\sigma^-\sigma^+\rho-\rho\sigma^-\sigma^+)\\&+\frac{P}{2}(1-p)(2\sigma^-\rho\sigma^+-\sigma^+\sigma^-\rho-\rho\sigma^+\sigma^-)\\&+\frac{P}{2}p(2\sigma^+\rho\sigma^--\sigma^-\sigma^+\rho-\rho\sigma^-\sigma^+),\end{aligned} \tag{10}$$

where the offset of the cavity frequency from the atomic frequency is $\Delta=\omega_A-\omega_c$. The quantities γ, Γ, and P are respectively the rate of field loss in the mirrors, the spontaneous emission rate, and the incoherent pump rate. The indicated dissipation constants can be expressed in terms of correlation functions of the corresponding reservoir operators $\{b_j\}$ and $\{\Pi_k\}$ (see, for example, Refs. 24 and 25).

The mean number of excitations of the atomic reservoir N_T in Eq. (10) in the case in which the reservoir $\{b_j\}$ is in thermodynamic equilibrium at temperature T is (see, e.g., Ref. 25)

$$\begin{aligned}N_T\equiv\overline{N}(\omega=\omega_A,T)&=\langle b_j^+b_j\rangle|_{\omega_j=\omega_A}\\&=[\exp(\hbar\omega_A/kT)-1]^{-1}.\end{aligned} \tag{11}$$

The interaction of an atom with the pump reservoir $\{\Pi_k\}$ leads to a deviation from thermal equilibrium, and the degree of excitation of the atom by the pump is determined by the mean number of Fermi quanta of the pump reservoir at the atomic transition frequency $p(\omega=\omega_A)$ in Eq. (10), for which we have

$$p=\langle\Pi_k^+\Pi_k\rangle|_{\omega_k=\omega_A}, \quad 0\leqslant p\leqslant 1. \tag{12}$$

The equation of motion for the averages of the atomic operators can be found from Eq. (10) with the help of the equation

$$\frac{\partial\langle\sigma^z\rangle}{\partial t}=\mathrm{Tr}\left(\frac{\partial\rho}{\partial t}\sigma^z\right). \tag{13}$$

Hence it follows that the mean population inversion of an atom interacting with the reservoir $\{b_j\}$ responsible for spontaneous decay obeys the equation

$$\frac{\partial\langle\sigma^z\rangle_{\mathrm{spont}}}{\partial t}=-(2N_T+1)\Gamma\langle\sigma^z\rangle_{\mathrm{spont}}-\Gamma. \tag{14}$$

The interaction with the pump reservoir $\{\Pi_k\}$, leading to a deviation from thermodynamic equilibrium, is described by the equation

$$\frac{\partial\langle\sigma^z\rangle_{\text{pump}}}{\partial t}=-P\langle\sigma^z\rangle_{\text{pump}}+(2p-1)P. \tag{15}$$

The solution of Eq. (14) under steady-state (SS) conditions has the form

$$\langle\sigma^z\rangle^{SS}_{\text{spont}}=-\frac{1}{2N_T+1}, \tag{16}$$

i.e., spontaneous decay of an excited state of an atom leads to a Boltzmann (Fermi) distribution of the populations if we set the thermal number of excitations equal to $N_T=[\exp(\hbar\omega_A/kT)-1]^{-1}$.

The steady-state solution of Eq. (15) shows that the action of the pump leads to steady-state inversion:

$$\langle\sigma^z\rangle^{SS}_{\text{spont}}=2p-1. \tag{17}$$

The parameter p thus characterizes the degree of steady-state excitation of the atomic states supported by the given pump in the absence of interaction with the field, $0\leqslant p\leqslant 1$.

Using Eq. (1) and the orthonormality of the basis of Fock states, as well as Eqs. (2), (5)–(7), and (10), we find a system of coupled differential equations for the elements of the density matrix having the following form:[26]

$$\begin{aligned}
&\rho_{1n,m}(t)\equiv\langle\uparrow|\rho_{n,m}|\uparrow\rangle+\langle\downarrow|\rho_{n,m}|\downarrow\rangle,\\
&\rho_{2n,m}(t)\equiv\langle\uparrow|\rho_{n,m}|\uparrow\rangle-\langle\downarrow|\rho_{n,m}|\downarrow\rangle,\\
&\rho_{3n,m}(t)\equiv\langle\uparrow|\rho_{n,m}|\downarrow\rangle,\quad \rho_{4n,m}(t)\equiv\langle\downarrow|\rho_{n,m}|\uparrow\rangle.
\end{aligned} \tag{18}$$

The system of equations for the matrix components (18) with dimension $4\times(n_{\max}+1)\times(n_{\max}+1)$ given in Appendix A was solved numerically. At the initial instant of time the field is in general in an arbitrary mixed state, and the atom is in a superposition of the upper and lower states. Thus, the density matrix of the atom and field, not interacting at the time $t=0$, is

$$\begin{aligned}
&\rho(0)=\rho_a\otimes\rho_f,\\
&\rho_a=\left[\cos\frac{\theta}{2}|\uparrow\rangle+\sin\frac{\theta}{2}e^{i\varphi}|\downarrow\rangle\right]\\
&\qquad\times\left[\cos\frac{\theta}{2}\langle\uparrow|+\sin\frac{\theta}{2}e^{-i\varphi}\langle\downarrow|\right],\\
&\rho_f=\sum_{n,m=0}^{\infty}c_nc_m^*|n\rangle\langle m|.
\end{aligned} \tag{19}$$

In particular, if the field at the initial time is in the coherent state $|\alpha\rangle$, then the relation for the coefficients of the expansion over Fock states

$$c_n^*c_m=\frac{\alpha^{*n}\alpha^m}{\sqrt{n!m!}}\exp(-|\alpha|^2).$$

If the initial state is a pure Fock state $|n_0\rangle$, then $c_n^*c_m=\delta_{n,m}$. In the thermal state we have $c_n^*c_m=[n_T^n/(1+n_T)^{n+1}]\delta_{n,m}$.

3. FLUCTUATIONS OF THE PHOTON NUMBER INSIDE THE CAVITY

The mean photon number, mean inversion, fluctuations (variance) of the photons, and the mean field, obtained by grouping matrix elements of a system of the form (18), are given by

$$\begin{aligned}
&\langle n(t)\rangle=\mathrm{Tr}(\rho(t)a^+a)=\sum_{n=0}^{\infty}n\rho_{1n,n}(t),\\
&\langle D(t)\rangle=\mathrm{Tr}(\rho(t)\sigma^z)=\sum_{n=0}^{\infty}\rho_{2n,n}(t),\\
&\mathrm{Var}(n(t))\equiv\langle(\Delta n(t))^2\rangle=\sum_{n=0}^{\infty}(n-\langle n(t)\rangle)^2\rho_{1n,n}(t),\\
&\langle a^+(t)\rangle=\sum_{n=0}^{\infty}\sqrt{n+1}\rho_{1n,n+1}(t),\\
&\langle a(t)\rangle=\sum_{n=1}^{\infty}\sqrt{n}\rho_{1n,n-1}(t),
\end{aligned} \tag{20}$$

The variances of the conjugate quadratures $X_+(t)=[a^+(t)+a(t)]/2$ and $X_-(t)=[a^+(t)-a(t)]/2i$ can be expressed in terms of the matrix elements of the density operator:

$$\begin{aligned}
\langle(\Delta X_\pm)^2\rangle=\frac{1}{4}\Bigg\{&\sum_{n=0}^{\infty}(2n+1)\rho_{1n,n}(t)\\
&\pm\sum_{n=2}^{\infty}\sqrt{n(n-1)}\rho_{1n,n-2}(t)\\
&\pm\sum_{n=0}^{\infty}\sqrt{(n+1)(n+2)}\rho_{1n,n+2}(t)\\
&\mp\left[\sum_{n=0}^{\infty}\sqrt{n+1}\rho_{1n,n+1}(t)\right.\\
&\left.\pm\sum_{n=1}^{\infty}\sqrt{n}\rho_{1n,n-1}(t)\right]^2\Bigg\}.
\end{aligned} \tag{21}$$

As the initial state of the field at $t=0$ we used both the coherent vacuum state and a random thermal state with mean photon number corresponding to the optical range $\omega_a\sim10^{14}\,\mathrm{s}^{-1}$ at $T\approx300\,\mathrm{K}$. A numerical study of systems (A3)–(A6) in the representation (18) for various values of the laser parameters showed that steady-state laser radiation possesses sub-Poissonian statistics of the photon number if the rate of spontaneous emission satisfies the condition $\Gamma<g,\gamma$. Maximum squeezing of the fluctuations of the photon number occurs for $\Gamma\ll g,\gamma$ for $T=0$ and amounts to 15% (see Fig. 1). With increasing $\Gamma\to\gamma$, the degree of squeezing decreases, and for $\Gamma>\gamma$ the radiation goes into the super-Poissonian state. The value of the pump parameter p has a substantial influence on the steady-state statistics of the radiation; the greatest squeezing occurs for $p=1$ and decreases for $p<1$. The existence of a frequency offset ($\Delta\neq0$) also has a negative effect on the degree of squeezing. The optimal value of the rate of cavity losses γ and of the pump rate P

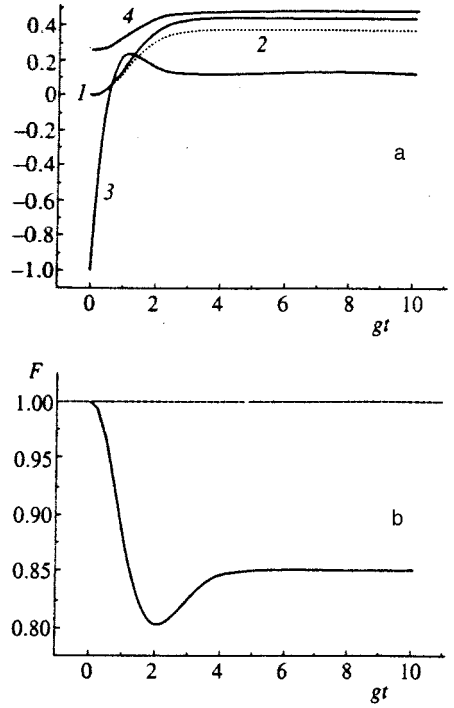

FIG. 1. a) Mean photon number $\langle n(t)\rangle$ (*1*), fluctuations (variance) of the photon number $\mathrm{Var}\, n(t)$ (*2*), mean inversion $\langle D(t)\rangle$ (*3*), and fluctuations (variance) of the field quadratures $\mathrm{Var}\, X_{\pm}(t)$ (*4*) of the laser for the initial state of the field in the coherent vacuum and of the atom in the lower level, for $\gamma=1.4g$, $\Gamma\ll\gamma$, $P=1.4g$, $p=1$, $\Delta=0$, and $T=0$. b) Fano factor $F=\mathrm{Var}\,(n)/\langle n\rangle$.

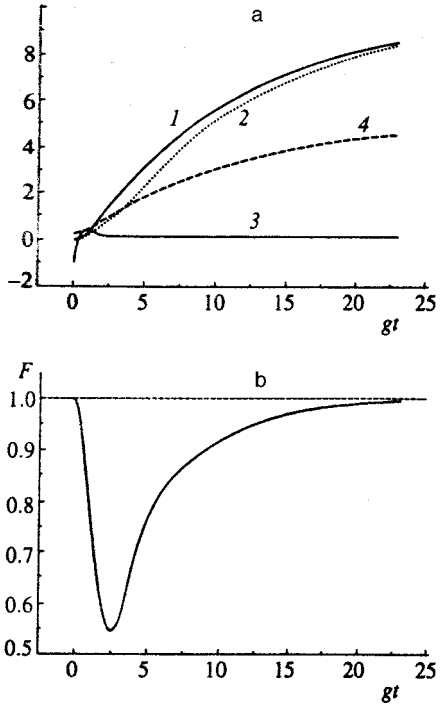

FIG. 2. The same as in Fig. 1, but for $\gamma=0.1g$, $\Gamma\ll\gamma$, $P=2g$, $p=1$, $\Delta=0$, and $T=0$.

relative to the coupling constant g are $\gamma=P\approx 1.4g$ for $p=1$, $\Delta=0$; here the Fano factor ($F=\mathrm{Var}(n)/\langle n\rangle$) is $F=0.85$ at temperatures $T<100$ K and $\omega_A=10^{14}$.

As can be seen from Fig. 2, as steady-state lasing builds up, the laser radiation remains in a state with nonclassical sub-Poissonian photon statistics. Reducing the rate of cavity losses relative to the pump rate makes it possible to dramatically reduce the Fano factor. Figure 2 depicts the dynamics of the Fano factor for $\gamma=0.1g$, $P=2g$ for the case of laser generation from the coherent vacuum state of the field with the atom in its lower level. In this case, the minimum value of the Fano factor $F_{\min}$ is 0.54 for $\langle n\rangle=1.43$. Thus, in the transient lasing regime, intense squeezed radiation can be generated. The degree of squeezing in the transient regime depends on the initial state. Maximum squeezing and intensity of the radiation are achieved if the atom is in the upper state and the field in the coherent vacuum state. In this case, for example, for $\gamma=0.1g$, $P=0.5g$ fivefold squeezing ($F_{\min}=0.19$) is achieved for $gt=1.8$ and $\langle n\rangle=1.2$ (Fig. 3).

Quadrature squeezing ($\mathrm{Var}X_{+}<1/4$ or $\mathrm{Var}X_{-}<1/4$) is absent for all of the laser parameter values we considered.

4. FLUCTUATIONS OF THE PHOTON NUMBER AT THE CAVITY OUTPUT

We assume that inside the laser cavity the electromagnetic field is in a state with discrete values of the frequencies (photon energies), whereas outside it the field has a continuous spectrum. As a consequence, it must be assumed that temporal fluctuations of the electromagnetic field inside the laser cavity are sources of fluctuations in the frequency spectrum of the radiation exiting through the resonator mirror. The field outside the cavity can be represented as a sum of the laser radiation field exiting through the mirror and the noise field of the reservoir–thermostat incident upon the mirror, i.e., $a^{(\mathrm{out})}(t)=b^{(\mathrm{in})}(t)+\sqrt{\gamma}\,a(t)$, where

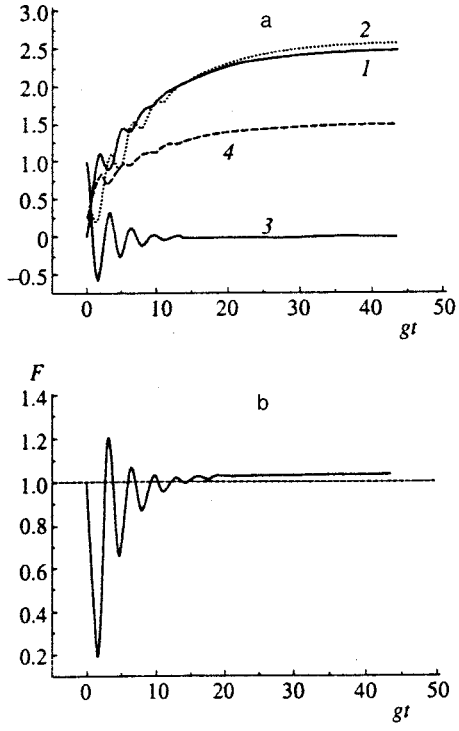

FIG. 3. The same as in Fig. 1, but for $\gamma=0.1g$, $\Gamma\ll\gamma$, $P=0.5g$, $p=1$, $\Delta=0$, and $T=0$, and the atom in the upper level at the initial time $t=0$.

$$b^{(\mathrm{in})}(t)\propto\sum_j \sqrt{\omega_j}\exp(-i\omega_j t)\,b_j$$

(Refs. 27–29). The Heisenberg operator $a^{(\mathrm{out})+}(t)\,a^{(\mathrm{out})}(t)$ is the photon number operator of photons exiting through the mirror per unit time. One quantity characterizing the statistics of the laser radiation passing through the output mirror is the steady-state fluctuation spectrum[12]

$$V^{(\mathrm{out})}(\omega)=\lim_{t\to\infty}2\int_0^\infty d\tau\cos\omega\tau[\langle n^{(\mathrm{out})}(t+\tau)n^{(\mathrm{out})}(t)\rangle - \langle a^{(\mathrm{out})+}(t+\tau)a^{(\mathrm{out})}(t)\rangle^2]. \quad (22)$$

Since the two-time correlators under steady-state conditions are even functions of τ, we have used the Fourier cosine transform in Eq. (22). The commutation relations for the field operators constituting the continuous spectrum outside the cavity have the form[27,28]

$$[a^{(\mathrm{out})}(t+\tau),a^{(\mathrm{out})+}(t)]=\delta(\tau). \quad (23)$$

Hence we find that the two-time correlation function of photon number operators is

$$\begin{aligned}&\langle n^{(\mathrm{out})}(t+\tau)\,n^{(\mathrm{out})}(t)\rangle\\ &\quad\equiv\langle a^+(t+\tau)\,a(t+\tau)\,a^+(t)\,a(t)\rangle^{(\mathrm{out})}\\ &\quad=\langle a^+(t+\tau)\,a(t)\rangle^{(\mathrm{out})}\delta(\tau)+\langle a^+(t)\,a^+(t+\tau)\,a\\ &\quad\times(t+\tau)\,a(t)\rangle^{(\mathrm{out})},\end{aligned} \quad (24)$$

i.e., the spectrum of fluctuations of the photon number of the field at the cavity output consists of terms in the shot noise and the chronologically and normal-ordered fluctuation spectrum:

$$V^{(\mathrm{out})}(\omega)=\langle n^{(\mathrm{out})}(t_{SS})\rangle+:V^{(\mathrm{out})}(\omega):. \quad (25)$$

In the case of one transparent mirror, the field correlators of the discrete mode inside the cavity are related to the field correlators of the continuous spectrum outside the cavity, as shown in Refs. 27–29:

$$\langle a^+(t)a(t)\rangle^{(\mathrm{out})}=\gamma\langle a^+(t)a(t)\rangle, \quad (26)$$

$$\begin{aligned}&\langle a^+(t)a^+(t+\tau)a(t+\tau)a(t)\rangle^{(\mathrm{out})}\\ &\quad=\gamma^2\langle a^+(t)a^+(t+\tau)a(t+\tau)a(t)\rangle.\end{aligned} \quad (27)$$

Substituting Eqs. (26) and (27) into Eq. (21), we finally obtain the following formula for the spectral Fano factor:

$$\begin{aligned}F(\omega)\equiv\frac{V^{(\mathrm{out})}(\omega)}{\langle n\rangle^{(\mathrm{out})}}&=1\\ &+\lim_{t\to\infty}\frac{2\gamma}{\langle a^+(t)a(t)\rangle}\int_0^\infty d\tau[\langle a^+(t)a^+(t+\tau)\\ &\times a(t+\tau)a(t)\rangle-\langle a^+(t+\tau)a(t+\tau)\rangle\\ &\times\langle a^+(t)a(t)\rangle]\cos\omega\tau.\end{aligned} \quad (28)$$

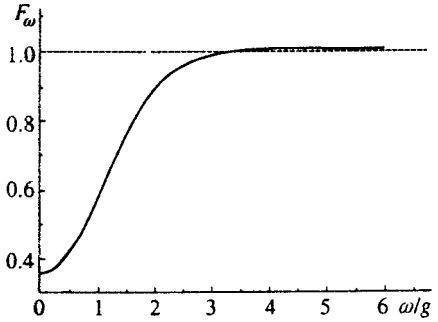

FIG. 4. Spectral Fano factor of laser radiation outside the cavity for pump parameters $P=1.4$, $p=1$, and losses $\gamma=2\gg\Gamma$ (in units of g).

Under steady-state conditions it is not hard to obtain an expression for the correlation functions of the field operators inside the cavity from the quantum regression theorem (see, e.g., Ref. 25)

$$\langle a^+(t)a^+(t+\tau)a(t+\tau)a(t)\rangle_{SS}=\mathrm{Tr}(a^+a\tilde{\rho}(\tau)), \quad (29)$$

where the operator $\tilde{\rho}(\tau)\equiv\tilde{\rho}(t+\tau)$ satisfies the Liouville equation (10) with initial condition ($\tau=0$)

$$\tilde{\rho}_{n,m}(0)=\sqrt{(n+1)(m+1)}\,\rho_{n+1,m+1}(t_{SS}). \quad (30)$$

Calculation shows that outside the cavity the squeezing of laser radiation is even more significant: for optimal values of the pump parameters ($P=1.4$, $p=1$) and loss parameter ($\gamma=2$), the spectral Fano factor is below the shot-noise limit by almost a factor of three and is 0.3567 at zero frequency (Fig. 4). The positions of optimal sqeezing outside and inside the cavity do not coincide: the largest squeezing occurs outside the cavity at a higher loss rate. As a consequence, the mean photon number of the radiation in this case is reduced: $\langle n\rangle=0.29$.

5. CONCLUSION

We have analyzed the dynamics of the quantum-statistical properties of the radiation of a one-atom laser. A significant deviation of the magnitude of the fluctuations of the photon number outside the cavity from the shot-noise limit (by up to 15%) is found to obtain under steady-state conditions, when the pump rate and the loss rate through the mirror are many times larger than the spontaneous decay rate and are comparable in magnitude to the coupling constant of an atomic electron with a cavity mode. We have shown that under special conditions, multiple squeezing is present in the transient dynamic regime. The squeezing of radiation exiting through the resonator mirror is significantly greater than its value inside the cavity: the Fano factor at zero frequency reaches 0.36.

The case we have considered, in which the coupling constant is much less than the spontaneous emission rate and is comparable in magnitude to the cavity loss and pump rates, is atypical of most known lasers. The indicated requirements are best satisfied by a laser that uses transitions between highly excited states of Rydberg atoms, where the coupling constant can reach $g\sim10^6\ \mathrm{s}^{-1}$.

In the present work we have shown that a one-atom two-level laser with incoherent pumping is a potential source of nonclassical radiation with fluctuations of the photon number significantly lower than the shot-noise limit.

APPENDIX A:

Using the relations for the creation and annihilation operators in the Fock basis

$$a|n\rangle=\sqrt{n}|n-1\rangle,\quad a^+|n\rangle=\sqrt{n+1}|n+1\rangle,$$
$$\langle n|a=\sqrt{n+1}\langle n+1|,\quad \langle n|a^+=\sqrt{n}\langle n-1|, \tag{A1}$$

Eqs. (2), (5)–(7), and the equalities

$$\sigma^z\sigma^+=\sigma^+,\quad \sigma^+\sigma^z=-\sigma^+,\quad \sigma^-\sigma^z=\sigma^-, \tag{A2}$$

we directly obtain from the Liouville equation (10) the equations of motion for the elements of the density matrix (9) in the form (18). As a result

$$\dot\rho_{1,n,m}=ig(\sqrt{m+1}\rho_{3,n,m+1}-\sqrt{n}\rho_{3,n-1,m}+\sqrt{m}\rho_{4,n,m-1}-\sqrt{n+1}\rho_{4,n+1,m})+\frac{\gamma}{2}L\rho_1, \tag{A3}$$

$$\dot\rho_{2,n,m}=ig(\sqrt{m+1}\rho_{3,n,m+1}+\sqrt{n}\rho_{3,n-1,m}-\sqrt{m}\rho_{4,n,m-1}-\sqrt{n+1}\rho_{4,n+1,m})+\frac{\gamma}{2}L\rho_2+[P(2p-1)-\Gamma]\rho_{1,n,m}-[\Gamma(2N_T+1)+P]\rho_{2,n,m}, \tag{A4}$$

$$\dot\rho_{3,n,m}=i\Delta\rho_{3,n,m}+i\frac{g}{2}(\sqrt{m}\rho_{1,n,m-1}-\sqrt{n+1}\rho_{1,n+1,m}+\sqrt{m}\rho_{2,n,m-1}+\sqrt{n+1}\rho_{2,n+1,m})+\frac{\gamma}{2}L\rho_3-\frac{\Gamma(2N_T+1)+P}{2}\rho_{3,n,m}, \tag{A5}$$

$$\dot\rho_{4,n,m}=i\Delta\rho_{4,n,m}+i\frac{g}{2}(\sqrt{m+1}\rho_{1,n,m+1}-\sqrt{n}\rho_{1,n-1,m}-\sqrt{m+1}\rho_{2,n,m+1}-\sqrt{n}\rho_{2,n-1,m})+\frac{\gamma}{2}L\rho_4-\frac{\Gamma(2N_T+1)+P}{2}\rho_{4,n,m}, \tag{A6}$$

where the term common to all four of these equations, due to cavity losses through the mirror, can be written

$$L\rho_j=(n_T+1)[2\sqrt{(n+1)(m+1)}\rho_{j,n+1,m+1}-(m+n)\rho_{j,n,m}]+n_T[2\sqrt{nm}\rho_{j,n-1,m-1}-\rho_{j,n,m}(n+m+2)]. \tag{A7}$$

*)E-mail: kozlovsky@neur.lpi.msk.su

[1] G. S. Agarwal and S. Dutta Gupta, Phys. Rev. A **42**, 1737 (1990).
[2] Y. Mu and C. M. Savage, Phys. Rev. A **46**, 5944 (1992).
[3] G. M. Meyer, H.-J. Briegel, and H. Walther, Europhys. Lett. **37**, 317 (1997).
[4] H.-J. Briegel, G. M. Meyer, and B.-G. Englert, Phys. Rev. A **53**, 1143 (1996).
[5] G. M. Meyer, M. Löffler, and H. Walther, Phys. Rev. A **56**, R1099 (1997).
[6] M. Löffler, G. M. Meyer, and H. Walther, Phys. Rev. A **55**, 3923 (1997).
[7] C. Ginzel, H.-J. Briegel, U. Martini, B.-G. Englert, and A. Schenzle, Phys. Rev. A **48**, 732 (1993).
[8] Yu. M. Golubev and I. V. Sokolov, Zh. Éksp. Teor. Fiz. **87**, 408 (1984) [Sov. Phys. JETP **60**, 234 (1984)].
[9] Y. Yamamoto, S. Mashida, and O. Nilsson, Phys. Rev. A **34**, 4025 (1986).
[10] C. Benkert, M. O. Scully, J. Bergou, L. Davidovich, M. Hillery, and M. Orszag, Phys. Rev. A **41**, 2756 (1990).
[11] C. Benkert and M. O. Scully, Phys. Rev. A **42**, 2817 (1990).
[12] L. Davidovich, J. Mod. Phys. **68**, 127 (1996).
[13] M. A. Marte, H. Ritsch, and D. F. Walls, Phys. Rev. Lett. **61**, 1093 (1988).
[14] T. A. B. Kennedy and D. F. Walls, Phys. Rev. A **40**, 6366 (1989).
[15] H. Ritsch, P. Zoller, C. W. Gardiner, and D. F. Walls, Phys. Rev. A **44**, 3361 (1991).
[16] H. Ritsch and P. Zoller, Phys. Rev. A **45**, 1881 (1992).
[17] T. C. Ralph and C. M. Savage, Phys. Rev. A **44**, 7809 (1991).
[18] D. L. Hart and T. A. B. Kennedy, Phys. Rev. A **44**, 4572 (1991).
[19] A. B. Kozlovskiĭ and A. N. Oraevskiĭ, Kvantovaya Elektron. **21**, 273 (1994).
[20] A. V. Kozlovskiĭ, Zh. Éksp. Teor. Fiz. **104**, 2995 (1993) [JETP **77**, 393 (1993)].
[21] A. Eschmann and C. W. Gardiner, Phys. Rev. A **54**, 3373 (1996).
[22] G. M. D. D'Ariano, C. Macchiavello, and M. G. A. Paris, Phys. Rev. Lett. **73**, 3187 (1994).
[23] A. B. Kozlovskiĭ and A. N. Oraevskiĭ, Zh. Éksp. Teor. Fiz. **109**, 1524 (1996) [JETP **82**, 820 (1996)].
[24] M. Lax, in *Fluctuations and Coherence Phenomena in Classical and Quantum Physics*, M. Chrétien, E. P. Gross, and S. Deser (eds.), Gordon and Breach, New York (1968).
[25] C. W. Gardiner, in *Handbook of Stochastic Methods*, H. Haken (ed.), Springer-Verlag, Berlin (1985).
[26] B. Daeubler, H. Risken, and J. Schoendorff, Phys. Rev. A **46**, 1654 (1992).
[27] C. W. Gardiner and M. J. Collett, Phys. Rev. A **31**, 3761 (1985).
[28] M. J. Collett and C. W. Gardiner, Phys. Rev. A **30**, 1386 (1984).
[29] H. J. Carmichael, J. Opt. Soc. Am. B **4**, 1588 (1987).

Translated by Paul F. Schippnick

Order $(Z\alpha)^6 m_e^2/m_\mu$ corrections to the fine structure of muonium

A. P. Martynenko*)

Samara State University, 443011 Samara, Russia

R. N. Faustov

Scientific Council ''Cybernetics,'' Russian Academy of Sciences, 101000 Moscow, Russia

(Submitted 19 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1221–1235 (April 1999)

We use the quasipotential method to calculate the total correction of order $(Z\alpha)^6 m_e^2/m_\mu$ in the energy spectrum of the $n\,^3S_1$ states of muonium. The numerical value of the muonium-fine-structure interval $2\,^3S_1-1\,^3S_1$ amounts to 0.19 MHz. © *1999 American Institute of Physics.* [S1063-7761(99)00504-1]

1. INTRODUCTION

The study of the fine structure of muonium and positronium is a check on the validity of electrodynamics that is sensitive to high-order radiative corrections in α (Ref. 1). Calculations of various contributions to the fine structure of the energy levels of a hydrogen-like system have been carried out by many researchers (see, e.g., Refs. 2–5) and interest in this problem is still very high.[6–9] Recent progress in calculating logarithmic contributions of the $\alpha^6\ln\alpha$ type in the positronium fine-structure intervals ($2\,^3S_1-1\,^3S_1$ and $2\,^3S_1-2\,^3P_J$; see Refs. 10–12) does not mean, however, that there is no need to calculate high-order corrections $O(\alpha^6)$ (Ref. 13). Corrections of order $(Z\alpha)^6 m_1^2/m_2$ to the s levels of the hydrogen atom were studied in Ref. 6–8 by various approaches to the bound-state problem, but as noted by Yelkhovsky,[9] there are still discrepancies among calculated corrections of the required order in the fine structure of a hydrogen-like atom. The development of experimental methods based on Doppler-free two-photon spectroscopy has made it possible to measure the ''large'' structure intervals in muonium and positronium:[14–16]

$$\Delta E^{\rm expt}_{\rm Ps}(2\,^3S_1-1\,^3S_1)=\begin{cases}1233\,607\,218.9\pm 10.7\ \text{MHz},\\ 1233\,607\,216.4\pm 3.2\ \text{MHz},\end{cases} \tag{1}$$

$$\Delta E^{\rm expt}_{\rm Mu}(2\,^3S_1-1\,^3S_1)=2455\,527\,936\pm 120\pm 140\ \text{MHz}. \tag{2}$$

The frequency of the Doppler-free two-photon transition $1S-2S$ in the hydrogen atom (as well as the hyperfine splitting of the ground state of the hydrogen atom) is a quantity that has been measured to high accuracy.[13] Since $\alpha^{-1}=137.035\,9979\,(32)$, the muon-to-electron mass ratio $m_\mu/m_e=206.768\,259\,(62)$, and for muonium the order $(Z\alpha)^6 m_e^2/m_\mu$ contribution may reach values of order 1 MHz. Hence the reduction to several megahertz of the experimental error in upcoming measurements[1)] of the fine structure of muonium (and positronium) makes calculations of the correction of the required order in α and m_e/m_μ highly desirable. In the present paper we study nuclear recoil corrections of order $(Z\alpha)^6 m_1/m_2$ in the fine structure of muonium. Bodwin *et al.*[17] calculated similar contributions to the hyperfine splitting of muonium.

In QED there are many approaches to describing the relativistic energy spectra of two-particle bound states.[2,3,13,18–20] The approaches differ in the way in which the calculations are organized, i.e., the type of equation used in describing the two-particle system, the way in which the particle interaction operator is set up, and the complexity of determining the corrections of required accuracy in the energy levels. However, all methods of describing the energy spectra yield equivalent results in a fixed perturbation-theory order in the small parameters α and m_1/m_2.

Our calculations are based on a local quasipotential equation of the Schrödinger type[21]

$$\left(\frac{b^2}{2\mu_R}-\frac{\mathbf{p}^2}{2\mu_R}\right)\psi_M(\mathbf{p})=\int\frac{d\mathbf{q}}{(2\pi)^3}V(\mathbf{p},\mathbf{q},M)\,\psi(\mathbf{q}), \tag{3}$$

where $b^2=E_1^2-m_1^2=E_2^2-m_2^2$, $\mu_R=E_1E_2/M$ is the relativistic reduced mass, $M=E_1+E_2$ is the mass of the bound state, and m_1 and m_2 are the electron and muon masses. For the initial approximation of the quasipotential $V(\mathbf{p},\mathbf{q},M)$ for the bound system $(e^-\mu^+)$ we take the ordinary Coulomb potential. In Ref. 22, using Eq. (3) as the staring point, the researchers determined the relativistic corrections $m\alpha^6$ in the fine structure of the positronium spectrum that are introduced by the one-photon interaction with allowance for the vertex corrections and corrections in the photon propagator, the two-photon exchange interaction, and second-order perturbation theory. The foremost among such corrections is

$$\Delta B_1=\frac{5m_1^2(Z\alpha)^6}{2m_2n^6}, \tag{4}$$

which is of order $(Z\alpha)^6 m_1/m_2$ and emerges because of the condition for quantization of the energy levels in the Coulomb interaction,

$$\frac{b^2}{\mu_R^2}=-\frac{\alpha^2}{n^2}, \tag{5}$$

transformed for the binding energy B of this system.

2. CONTRIBUTION OF THE ONE-PHOTON INTERACTION TO THE FINE STRUCTURE

One-photon interaction provides the principal contribution to the energy spectrum of the two-particle bound state. The method of setting up a quasipotential for one-photon interaction in a system of two spinor particles was thoroughly studied in Refs. 22 and 23. Note that here it is convenient to use the relativistic projection operator on state $^3S_1$,

$$\hat{\Pi}=\frac{1}{2\sqrt{2}}\frac{(\hat{p}_1+m_1)(1+\gamma_0)\hat{\varepsilon}(-\hat{p}_2+m_2)}{\sqrt{\varepsilon_1+m_1}\sqrt{\varepsilon_2+m_2}}, \tag{6}$$

where p_1 and p_2 are the 4-momenta of the electron and muon in the initial state, and ε^μ is the muonium polarization vector. Expanding all relativistic factors in powers of the two small parameters $|\mathbf{p}|/m_{1,2}$ ($|\mathbf{p}|\sim Z\alpha$ is the momentum of the relative motion of the particles) and m_1/m_2 (the electron-to-muon mass ratio) so as to isolate the contribution of sixth order in $Z\alpha$ and first order in m_1/m_2, we can write the expression for the particle interaction operator in the coordinate representation in the form

$$\begin{aligned}V_1(r)&=-\frac{Z\alpha}{r}-\frac{\mu_R(Z\alpha)^2}{2m_1^2r^2}\left(1+\frac{2m_1}{m_2}\right)-\frac{Z\alpha}{4m_1^2r^3}(\mathbf{r}\cdot\boldsymbol{\nabla})\\&\times\left(1+\frac{4m_1}{m_2}\right)-\frac{\pi Z\alpha}{3m_1m_2}\delta\mathbf{r}-\frac{Z\alpha b^2}{m_1m_2r}\\&=V_c(r)+\Delta V_1(r).\end{aligned} \tag{7}$$

Note that the part ΔV_1 of the quasipotential (7), combined with quantization condition (5), correctly reproduces the known energy spectrum of the S states of muonium to within $O(\alpha^4)$ (Refs. 13 and 20). In our approach, the terms in ΔV_1 also provide order $O(\alpha^6)$ corrections to the energy spectrum, which is due to the dependence of the relativistic reduced mass μ_R and of b^2 on α. Averaging (7) over Coulomb wave functions[24] and extracting the required terms, we find that

$$\Delta B_2=-\frac{3m_1^2(Z\alpha)^6}{4m_2n^5}\left(5+\frac{2}{n}\right). \tag{8}$$

The one-photon interaction quasipotential also contains other terms that lead to order $(Z\alpha)^6$ corrections in the energy spectrum. These terms are obtained by setting up $V_{1\gamma}$ to within $|\mathbf{p}|^2/m_{1,2}^4$ and $|\mathbf{q}|^4/m_{1,2}^4$, and take the form

$$\begin{aligned}&\Delta V_2(\mathbf{p},\mathbf{q},M)\\&=-\frac{4\pi Z\alpha}{\mathbf{k}^2}\left\{\frac{b^4}{16m_1^4}\left(3-\frac{2m_1}{m_2}\right)+\frac{\mathbf{p}^4+\mathbf{q}^4}{96m_1^4}\right.\\&\times\left(3+\frac{m_1}{m_2}\right)-\frac{(\mathbf{p}^2+\mathbf{q}^2)(\mathbf{p}\cdot\mathbf{q})}{96m_1^4}\left(6+\frac{13m_1}{m_2}\right)\\&\left.-\frac{(\mathbf{p}^2+\mathbf{q}^2)b^2}{96m_1^4}\left(3-\frac{m_1}{m_2}\right)-\frac{(\mathbf{p}\cdot\mathbf{q})b^2}{48m_1^4}\left(3+\frac{7m_1}{m_2}\right)\right\}.\end{aligned} \tag{9}$$

Now we turn to calculations of order $(Z\alpha)^6$ corrections that originate in the terms in ΔV_2. We note that some terms in ΔV_2 lead to divergent integrals in the energy spectrum. The reason for this divergence lies in the power series expansion in $|\mathbf{p}|/m_{1,2}$ used in setting up ΔV_2. A typical divergent integral is $\int\mathbf{p}^2\,d\mathbf{p}\,\psi_{nS}(\mathbf{p})$. The relativistic order α^6 correction in this case is determined by the residue of the integrand at the pole of the wave function $\psi_{nS}(\mathbf{p})$ (Ref. 18). The calculation of this integral for an arbitrary value of the principal quantum number n yields

$$\begin{aligned}\int\frac{d\mathbf{p}}{(2\pi)^3}\frac{\mathbf{p}^2}{\mu_R^2}\psi_{nS}(\mathbf{p})&=-\frac{[3+2(n-1)(n+1)]}{n^2}\\&\times\alpha^2\psi_{nS}(\mathbf{r}=0).\end{aligned} \tag{10}$$

Using (10) in averaging of ΔV_2, we calculated the relativistic corrections of the proper order for levels with arbitrary values of the principal quantum number n:

$$\begin{aligned}\Delta B_3=\frac{m_1^2(Z\alpha)^6}{m_2}&\left(-\frac{139}{72n^3}+\frac{17}{12}\frac{\ln 2}{n^3}+\frac{73}{72}\frac{1}{n^5}+\frac{43}{96}\frac{1}{n^6}\right.\\&\left.+\frac{17}{12}(-1)^n\frac{1}{n^3}[C+\psi(n)-1]\right),\end{aligned} \tag{11}$$

where $\psi(z)=d\ln\Gamma(z)/dz$, and $C=0.577\,215\,6649\ldots$ is Euler's constant.

3. SECOND-ORDER PERTURBATION THEORY

In our case the second-order perturbation-theory correction in the muonium energy spectrum is determined by the sum of two terms:[25]

$$\begin{aligned}\Delta B^{(2)}&=\langle\psi_n^c|\Delta V_1|\psi_n^c\rangle\langle\psi_n^c|\frac{\partial\Delta V_1}{\partial B}|\psi_n^c\rangle\\&+\sum_{k=1,k\neq n}^{\infty}\frac{\langle\psi_n^c|\Delta V_1|\psi_k^c\rangle\langle\psi_k^c|\Delta V_1|\psi_k^c\rangle}{B_n^c-B_k^c}.\end{aligned} \tag{12}$$

The quasipotential (7) depends explicitly on the binding energy B (factors b^2 and μ_R). Bearing in mind that $\partial b^2/\partial B=2\mu$ to the required accuracy, we can write the contribution of the first term on the right-hand side of (12) to the energy spectrum as follows:

$$\Delta B_4=\langle\psi_n^c|\Delta V_1|\psi_n^c\rangle\langle\psi_n^c|\frac{\partial\Delta V_1}{\partial B}|\psi_n^c\rangle=\frac{m_1^2}{m_2}(Z\alpha)^6\frac{1}{n^5}. \tag{13}$$

The second term on the right-hand side of (12) is determined by the reduced nonrelativistic Coulomb Green's function,[19,26–31] whose partial expansion is

$$\overline{G}_n(\mathbf{r},\mathbf{r}',B)=\sum_{l,m}\overline{g}_{nl}(r,r',B)Y_{lm}(\mathbf{n})Y_{lm}^*(\mathbf{n}'). \tag{14}$$

TABLE I. Second-order perturbation-theory contributions determined by the RCGF and the quasipotential (7) (in units of $(Z\alpha)^6 m_1^2/m_2$).

ΔV_1	$-\frac{Z\alpha b^2}{m_1 m_2 r}$	$-\frac{(Z\alpha)^2\mu_R}{2r^2 m_1^2}\left(1+\frac{2m_1}{m_2}\right)$	$-\frac{Z\alpha(\mathbf{r}\cdot\boldsymbol{\nabla})}{4m_1^2 r^3}\left(1+\frac{4m_1}{m_2}\right)$	$-\frac{\pi Z\alpha}{3m_1 m_2}\delta(\mathbf{r})$
$-\frac{Z\alpha b^2}{m_1 m_2 r}$	–	$\frac{2-n}{n^5}$	$\frac{2n^2-5n+1}{4n^6}$	–
$-\frac{(Z\alpha)^2\mu_R}{2m_1^2 r^2}\left(1+\frac{2m_1}{m_2}\right)$	$\frac{2-n}{n^5}$	$\frac{2n+3}{2n^4}$	$-\frac{n^2+3n-1}{4n^5}$	$-\frac{4n+9}{6n^4}$
$-\frac{Z\alpha(\mathbf{r}\cdot\boldsymbol{\nabla})}{4m_1^2 r^3}\left(1+\frac{4m_1}{m_2}\right)$	$\frac{2n^2-5n+1}{4n^6}$	$-\frac{n^2+3n-1}{4n^5}$	$-\frac{n(n-1)(n+1)}{24n^6}$	$-\frac{n^2-6n+8}{12n^5}$
$-\frac{\pi Z\alpha}{3m_1 m_2}\delta(\mathbf{r})$	–	$-\frac{4n+9}{6n^4}$	$-\frac{n^2-6n+8}{12n^5}$	–

The radial function $\overline{g}_{nl}(r,r',B)$ was derived in Ref. 31 as a Sturm expansion in Laguerre polynomials. For the S state this function is

$$\overline{g}_{n0}(r,r',B_n) = -\frac{4Z\alpha\mu^2}{n}\left[e^{-(x+x')/2}\sum_{m=1,m\neq n}^{\infty}\frac{L^1_{m-1}(x)L^1_{m-1}(x')}{m(m-n)} + \frac{1}{n^2}\left(\frac{5}{2}+x\frac{\partial}{\partial x}+x'\frac{\partial}{\partial x'}\right)e^{-(x+x')/2}L^1_{n-1}(x)L^1_{n-1}(x')\right], \tag{15}$$

where $x=2\mu Z\alpha r/n$, and L_n^m are the ordinary Laguerre polynomials, defined by

$$L_n^m(x)=\frac{e^x x^{-m}}{n!}\left(\frac{d}{dx}\right)^n(e^{-x}x^{n+m}). \tag{16}$$

The reduced Coulomb Green's function (RCGF) (15) depends on two variables, r and r', but in calculating the corrections in (12) of a delta-function potential we must know the RCGF at $\mathbf{r}=0$. An expression for the RCGF in this case can be obtained via the Hostler representation for the Coulomb Green's function (see Ref. 32) by subtracting the pole term:

$$\overline{G}_n(\mathbf{r},0,B_n)=-\frac{Z\alpha\mu^2}{n\pi x}e^{-x/2}\sum_{s=0}^{n-1}\frac{(-x)^{n-s}}{s!}\frac{n!}{[(n-s)!]^2}\times\left\{(n-s)\left[\psi(n+1)-2\psi(n-s+1)-\frac{2(n-s)+3-x}{2n}+\ln x\right]+1\right\}. \tag{17}$$

In contrast to Ref. 32, this formula does not contain the free two-particle Green's function $G^f(r)=-\mu_R e^{-Z\alpha\mu_R r}/2\pi r$, which determines the iterative part of the quasipotential. The contribution of this function will be obtained separately.

As an example, we calculate the energy corrections in second-order perturbation theory that are determined by the delta-function potential and the term $\Delta V_1\propto 1/r^2$. This contribution can be written

$$\delta B=-\frac{\mu^5(Z\alpha)^6}{3m_1 m_2 n^4}\sum_{k=1}^{n}(-1)^k\frac{n!}{(n-k)!\,(k!)^2}\times\int_0^\infty x^{k-1}e^{-x}L^1_{n-1}(x)\,dx\left\{k\left[\psi(n+1)-2\psi(k+1)-\frac{2k+3-x}{2n}+\ln x\right]+1\right\}. \tag{18}$$

The expression (18) contains integrals of two types, with power-law and logarithmic functions, respectively. Evaluating the first integral with respect to the variable x,

$$I_1=\int_0^\infty x^{k-1}e^{-x}L^1_{n-1}(x)\,dx=\frac{(k-1)!\,\Gamma(n+1-k)}{(n-1)!\,\Gamma(2-k)}, \tag{19}$$

we see that only the term with $k=1$ remains in the sum. The second integral in (18),

$$I_2=\int_0^\infty x^{k-1}\ln x e^{-x}L^1_{n-1}(x)\,dx=\frac{(2-k)_{n-1}\Gamma(k)}{(n-1)!}[\psi(k)+\psi(2-k)-\psi(n+1-k)], \tag{20}$$

leads to a sum of the type

$$\sum_{k=1}^{n}(-1)^k\frac{\psi(2-k)(2-k)_{n-1}}{(n-k)!\,k!}=-\frac{n-1}{n}+C, \tag{21}$$

where

$$C=\lim_{n\to\infty}\left[-\ln n+\sum_{m=1}^{n}\frac{1}{m}\right]=0.577\,215\,66\ldots$$

is Euler's constant. If we now allow for the fact that

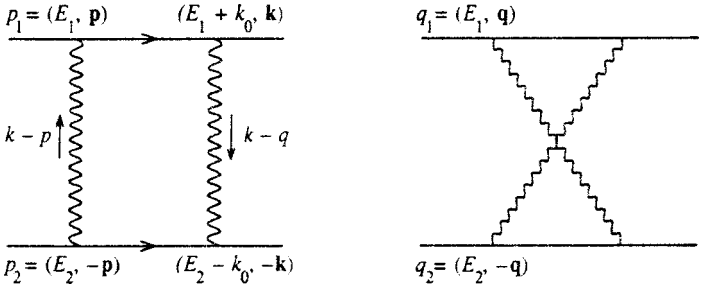

FIG. 1. Direct and crossed diagrams of two-photon exchange interaction.

$$\lim_{k\to n}\frac{\Gamma(n-k)}{\Gamma(1-k)}=(-1)^{n-1}(n-1)!, \tag{22}$$

we finally obtain an expression for the correction (18):

$$\delta B=-\frac{m_1^2}{6m_2}(Z\alpha)^6\frac{1}{n^4}(4n+9). \tag{23}$$

Reasoning along similar lines, we can calculate the contributions of the other terms in the quasipotential (7) in second-order perturbation theory via (12), (15), and (17). The results of such calculations are listed in Table I. The first column and the first row contain the various terms in ΔV_1. The value of the second integral in (12), with the various terms in ΔV_1 expressed in units of $(Z\alpha)^6 m_1^2/m_2$, can be found at the intersection of the appropriate row and column. The total contribution of the RCGF to the energy spectrum (without G^f) is

$$\Delta B_5=\left(-\frac{25}{24}-\frac{3}{n}-\frac{49}{24n^2}+\frac{3}{2n^3}\right)\frac{m_1^2(Z\alpha)^6}{m_2 n^3}. \tag{24}$$

We now examine the contribution of the free two-particle propagator to the correction $\Delta B^{(2)}$ [this contribution was ignored in the RCGF given by (17)]. It is convenient to do this in the momentum representation. Bearing in mind that

$$G^f(\mathbf{p},\mathbf{q},M)=\frac{(2\pi)^3\delta(\mathbf{p}-\mathbf{q})}{b^2/2\mu_R-\mathbf{p}^2/2\mu_R} \tag{25}$$

and that the delta-function term in the quasipotential (7) already contains the muon mass in the denominator, we can write the required iterative correction as

$$\Delta B_6=\frac{2\mu(Z\alpha)^2\pi^2\psi_{nS}(0)}{3m_2m_1^3}\int\frac{d\mathbf{p}}{(2\pi)^3}\psi_{nS}(\mathbf{p}) \times\left[\frac{\mathbf{p}\cdot\mathbf{k}}{\mathbf{k}^2}-\frac{\mathbf{p}^2+W^2}{\mathbf{k}^2}\right]\frac{d\mathbf{q}}{(2\pi)^3(\mathbf{q}^2+W^2)}, \tag{26}$$

where $W^2=\mu_R^2(Z\alpha)^2/n^2$. The divergence of this integral is the same as in (10). Using Feynman's parametrization and formula (10) in calculating (26), we find that

$$\Delta B_6=-\frac{m_1^2(Z\alpha)^6}{12m_2}\left(\frac{1}{n^3}-\frac{2\ln 2}{n^3}-\frac{2}{n^4} -\frac{2}{n^3}(-1)^n[C+\psi(n)-1]\right). \tag{27}$$

4. TWO-PHOTON EXCHANGE INTERACTION

The two-photon exchange interaction amplitude is represented by the two diagrams in Fig. 1. The particle interaction operators corresponding to these diagrams are

$$V_{2\gamma}^{(a)}(\mathbf{p},\mathbf{q})=\frac{i(Z\alpha)^2}{\pi^2} \times\int\frac{f_1(k,m_1,m_2)\,d^4k}{[(k-p)^2+i\epsilon][(k-q)^2+i\epsilon]D_e(k)D_\mu(-k)}, \tag{28}$$

$$f_1(k,m_1,m_2)=m_2(4m_1+2k_0)-2m_1k_0-2k_0^2+\tfrac{2}{3}k^2,$$

$$D_\mu(-k)=k^2-2E_2k_0+b^2+i\epsilon\approx-2m_2k_0+i\epsilon, \tag{28}$$

$$V_{2\gamma}^{(b)}(\mathbf{p},\mathbf{q})=\frac{i(Z\alpha)^2}{\pi^2} \times\int\frac{f_2(k,m_1,m_2)\,d^4k}{[(k-p)^2+i\epsilon][(k-q)^2+i\epsilon]D_e(k)D_\mu(p-q-k)},$$

$$f_2(k,m_1,m_2)=m_2(4m_1+2k_0)-2m_1k_0-6k_0^2 +\tfrac{10}{3}\mathbf{p}\cdot\mathbf{k}+\tfrac{10}{3}\mathbf{q}\cdot\mathbf{k}+\tfrac{10}{3}k^2,$$

$$D_\mu(p-q-k)=k^2+2E_2k_0+2\mathbf{k}\cdot(\mathbf{p}+\mathbf{q})-(\mathbf{p}+\mathbf{q})^2 +b^2+i\epsilon\approx2m_2k_0+i\epsilon. \tag{29}$$

The principal contribution of $V_{2\gamma}$ to the energy spectrum is proportional to α^5. Order $(Z\alpha)^6$ corrections can also appear in the energy levels if we allow, e.g., for the contribution of photon poles, whereupon $k_0\sim\alpha$, $|\mathbf{p}|\sim\alpha$, $|\mathbf{q}|\sim\alpha$, and $|\mathbf{k}|\sim\alpha$. To separate these terms, we transform the product of the electron and muon denominators in the direct two-photon diagram as follows:

$$\frac{1}{D_e(k)D_\mu(-k)}=\frac{-2\pi i\delta(k_0)}{-2E(\mathbf{k}^2-b^2)}-\frac{1}{2E}\left[\frac{1}{(k_0+i\epsilon)D_e(k)} +\frac{1}{(-k_0+i\epsilon)D_\mu(-k)}\right], \tag{30}$$

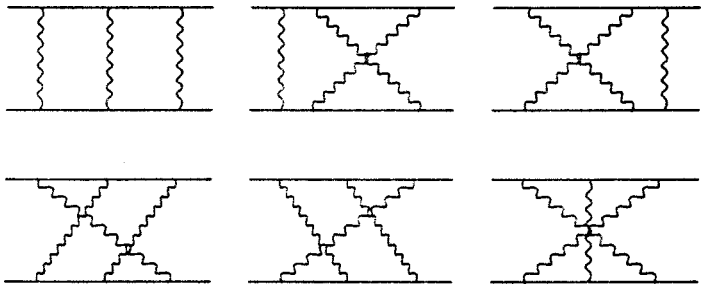

FIG. 2. Feynman diagrams of three-photon exchange interaction in the $(e^-\mu^+)$ system.

where the first term $[\sim\delta(k_0)]$ on the right-hand side and the iterative term of the quasipotential cancel out. The first term in the square brackets has the same structure in the leading order in $1/m_2$ as the crossed amplitude. As a result, the two-photon interaction quasipotential, which leads to the required correction of order $(Z\alpha)^6 m_1^2/m_2$, is

$$V_{2\gamma}(\mathbf{p},\mathbf{q}) = \frac{2i(Z\alpha)^2}{3\pi^2} \times \int \frac{d^4k\,[4\mathbf{k}^2+5\mathbf{k}\cdot(\mathbf{p}+\mathbf{q})-6k_0^2]}{[(k-p)^2+i\epsilon][(k-q)^2+i\epsilon]D_e(k)(2m_2k_0+i\epsilon)}. \tag{31}$$

The contribution of $V_{2\gamma}$ to the energy spectrum was calculated separately for $n=1$ and $n=2$ using *Mathematica* (Ref. 33) (the feynpar.m package), and the results of these calculations are

$$\Delta B_{2\gamma} = \begin{cases} -\dfrac{7m_1^2}{2m_2}(Z\alpha)^6, & n=1, \\ -\dfrac{31m_1^2}{16m_2}(Z\alpha)^6, & n=2. \end{cases} \tag{32}$$

5. THREE-PHOTON EXCHANGE INTERACTION

There are six diagrams, depicted in Fig. 2, that determine the amplitude of three-photon exchange interaction in muonium.

In the first diagram, the corresponding amplitude contains the factor α^6, which emerges due to the electromagnetic interaction vertices and the Coulomb wave functions. Hence, in the first stage of calculations we ignore the momentum vectors due to the relative motion of the electron and muon in the initial and final states, bearing in mind that the required numerical accuracy is already ensured. Then the amplitude representing the first diagram in Fig. 2 is

$$T_1^{2\gamma} = -\frac{(Z\alpha)^3}{4\pi^5}\int d^4p \int d^4p' \times \frac{\langle \gamma_1^\lambda(\hat q_1-\hat p'+m_1)\gamma_1^\nu(\hat p_1-\hat p+m_1)\gamma_1^\mu\rangle}{(p^2-w^2+i\epsilon)(p'^2-w^2+i\epsilon)[(p-p')^2+i\epsilon]} \times \frac{\langle \gamma_2^\mu(\hat p_2+\hat p+m_2)\gamma_2^\nu(\hat q_2+\hat p'+m_2)\gamma_2^\lambda\rangle}{D_e(p)D_e(p')D_\mu(-p)D_\mu(-p')}, \tag{33}$$

where $D_{e,\mu}(p)$ are the denominators of the electron and muon propagators,

$$D(\pm p) = p^2-w^2\pm 2mp^0+i\epsilon, \quad w^2=-b^2, \tag{34}$$

and the angle brackets indicate averaging over the Dirac bispinors; p_1 and p_2 are the 4-momenta of the particles in the initial state and q_1 and q_2, in the final state. As usual, the factor $Z\alpha$ emphasizes the exchange nature of the photon interaction of the particles.

The propagators of the exchange photons were chosen in the covariant Feynman gauge. As is known, the Coulomb gauge is the most natural one for exchange photons, since the Coulomb interaction is predominant in the $(e^-\mu^+)$ system. Nevertheless, Bodwin et al.[17] demonstrated that the Coulomb and Feynman gauges are equivalent in calculations of three-photon diagrams in the scattering approximation. To construct the quasipotential from the amplitude $T_1^{3\gamma}$ describing the interaction in the $(e^-\mu^+)$ system with $L=0$ and $J=1$, we introduce in the initial and final states the projection operator (6) and assume, in addition, that $\mathbf{p}=\mathbf{q}=0$. The use of (6) makes it possible to avoid cumbersome matrix multiplication in bispinor contractions and to proceed immediately with the calculation of the general trace in (33). As a result, the quasipotential of the first exchange diagram can be written

$$V_1^{3\gamma} = -\frac{(Z\alpha)^3}{\pi^5}\int d^4p \int d^4p' \times \frac{F_1(p,p')}{D_\gamma(p)D_\gamma(p')D_\gamma(p-p')D_e(p)D_e(p')D_\mu(-p)D_\mu(-p')}, \quad D_\gamma(p)=p^2-w^2+i\epsilon, \tag{35}$$

where

$$F_1(p,p')=f_{12}(p,p')m_2^2+\tfrac{1}{3}f_{11}m_2,$$
$$f_{12}=pp'-4m_1^2-2m_1p_0-2m_1p_0'-2p_0p_0',$$
$$f_{11}(p,p')=2m_1p'^2+p_0p'^2+10m_1pp'+2p_0pp'+2p_0'pp'+2m_1p^2+p_0'p^2+6m_1^2p_0+6m_1^2p_0'+4m_1p_0^2+4m_1p_0'^2-4m_1p_0p_0'. \quad (36)$$

The only terms that we kept in (36) were those proportional to m_2^2 and m_2, since we had in mind obtaining the contribution to the muonium fine structure only in the leading order in the parameter m_1/m_2. Below we find that we cannot limit ourselves in $F(p,p')$ to terms proportional to m_2^2. The quasipotentials of the other five amplitudes can be set up in a similar way. They differ from each other in the dependence on the momentum arguments in the muon denominators and in the type of function f_{i1} $(i=1,\ldots,6)$. The terms in $F_i(p,p')$ proportional to m_2^2 are the same in all six amplitudes. Note that if $\hat{\varepsilon}$ is replaced by γ_5 in the projection operator (6) (the $^1S_0$ state), we obtain the same function $f_{12}(p,p')$ as for the $^3S_1$ state of muonium. This means that the hyperfine splitting in muonium emerges as a higher-order effect in m_1/m_2. The functions f_{i1} are given by the following formulas:

$$f_{21}=-10m_1p'^2-5p_0p'^2+10m_1pp'+4p_0pp'-4p_0'pp'+2m_1p^2+4p_0'p^2+12p_0m_1^2-6m_1^2p_0'+4m_1p_0^2+8m_1p_0p_0'-8m_1p_0'^2+4p_0'p_0^2, \quad (37)$$

$$f_{31}=2m_1^2p'^2+4p_0p'^2+10m_1pp'-4p_0pp'+4p_0'pp'-10m_1p^2-5p_0'p^2-6m_1^2p_0+12m_1^2p_0'-8m_1p_0^2+8m_1p_0p_0'+4m_1p_0'^2+4p_0p_0'^2, \quad (38)$$

$$f_{41}=2m_1p'^2+p_0p'^2-2m_1pp'+4p_0pp'+2p_0'pp'-10m_1p^2-8p_0'p^2-12m_1^2p_0+6m_1^2p_0'+4m_1p_0^2-4m_1p_0p_0'+4m_1p_0'^2-8p_0'p_0^2, \quad (39)$$

$$f_{51}=-10m_1p'^2-8p_0p'^2-2m_1pp'+2p_0pp'+4p_0'pp'+2m_1p^2+p_0'p^2+6m_1^2p_0-12m_1^2p_0'+4m_1p_0^2-4m_1p_0p_0'+4m_1p_0'^2-8p_0'^2p_0, \quad (40)$$

$$f_{61}=-10m_1p'^2-5p_0p'^2-2m_1pp'-4p_0pp'-4p_0'pp'-10m_1p^2-5p_0'p^2-6m_1^2p_0'-8m_1p_0^2-4m_1p_0p_0'-8m_1p_0'^2. \quad (41)$$

The integrand in (33) has simple poles in variables p_0 and p_0' in the electron, muon, and photon propagators. Hence, the most natural way to integrate in (31) is to evaluate the integrals with respect to the energies p_0 and p_0' in the initial state via residue theory. Nevertheless, this method leads to extremely complicated intermediate expressions, which makes subsequent analytic integration over the spatial momenta $\mathbf{p}$ and $\mathbf{p}'$ highly problematic. We have therefore taken a different approach to integration in (33), transforming the denominators of the muon propagators with an eye to achieving the required numerical accuracy in m_1/m_2. Assuming that the spatial momentum $|\mathbf{p}|$ of the muon in the intermediate state is less than m_2, we find that

$$D_\mu(p)=p^2-w^2+2m_2p_0\approx 2m_2\left(p_0-\frac{\mathbf{p}^2+w^2}{2m_2}+i\epsilon\right)\approx 2m_2(p_0+i\epsilon), \quad (42)$$

where the second approximate equality means that we have ignored the muon kinetic energy in the intermediate state. Here we assume that the path of integration with respect to p_0 is closed in the lower half-plane. If in the numerators of the six amplitudes we examine terms proportional to m_2^2 [the functions $f_{12}(p,p')$], we see that we must transform the sum of terms with muon denominators in (30). If we then use the second approximate equality in (42), we find that

$$\frac{1}{D_\mu(-p)D_\mu(-p')}+\frac{1}{D_\mu(-p)D_\mu(p'-p)}+\frac{1}{D_\mu(-p')D_\mu(p-p')}+\frac{1}{D_\mu(p)D_\mu(p-p')}+\frac{1}{D_\mu(p')D_\mu(p'-p)}+\frac{1}{D_\mu(p')D_\mu(p)}\approx\frac{(-2\pi i)\,\delta(p_0)}{2m_2}\,\frac{(-2\pi i)\,\delta(p_0')}{2m_2}. \quad (43)$$

In the energy spectrum these terms lead to order α^4 corrections, which cancel similar terms from the iterative terms in the quasipotential. Hence to calculate the order α^6 contributions of interest, we must use the first approximate equality in (42). Now we take the difference

$$\frac{1}{2m_2(p_0-(\mathbf{p}^2+w^2)/2m_2+i\epsilon)}-\frac{1}{2m_2(p_0+i\epsilon)}\approx\frac{\mathbf{p}^2+w^2}{4m_2^2(p_0+i\epsilon)^2}, \quad (44)$$

and write $1/D_\mu(p)$ in the form

$$\frac{1}{D_\mu(p)}\approx\frac{1}{2m_2(p_0+i\epsilon)}+\frac{\mathbf{p}^2+w^2}{4m_2^2(p_0+i\epsilon)^2}. \quad (45)$$

The second term on the right-hand side of (41) is of a higher order in m_1/m_2 than the first, but provides a correction of the required order in α. Using the representation (45), from (43) we extract the terms of the required order in α. These terms are

TABLE II. Two-loop integrals K_i of the type (48) in the three-photon exchange integrals emerging in calculations of muonium fine structure.

	$\mathbf{p}^2(\mathbf{p}\cdot\mathbf{p}')$	$(\mathbf{p}\cdot\mathbf{p}')^2$	$\mathbf{p}'^2(\mathbf{p}\cdot\mathbf{p}')$	$w^2(\mathbf{p}\cdot\mathbf{p}')$
K_1	$2\ln 2-\frac{1}{2}$	$2\ln 2-\frac{1}{2}$	$2\ln 2-\frac{1}{2}$	0
	$\mathbf{p}^2(\mathbf{p}'^2-\mathbf{p}\cdot\mathbf{p}')$	$\mathbf{p}\cdot\mathbf{p}'(\mathbf{p}'^2-\mathbf{p}\cdot\mathbf{p}')$	$w^2(\mathbf{p}'^2-\mathbf{p}\cdot\mathbf{p}')$	$w^2\mathbf{p}^2$
K_2	$\frac{1}{2}\ln\frac{m_1}{2w}-\frac{1}{32}$	$\frac{1}{4}\ln\frac{m_1}{2w}-\frac{13}{32}$	$\frac{5}{32}$	$\frac{2}{3}$
	$\mathbf{p}'^2$	$\mathbf{p}\mathbf{p}'$	$\mathbf{p}^2$	w^2
K_3	–	$\frac{1}{4}\ln\frac{m_1}{2w}-\frac{1}{4}$	$\frac{1}{2}\ln\frac{m_1}{2w}-\frac{1}{8}$	$\frac{1}{8}$
K_4	$\frac{1}{2}\ln\frac{m_1}{2w}-\frac{1}{8}$	$\frac{1}{4}\ln\frac{m_1}{2w}-\frac{1}{4}$	–	$\frac{1}{8}$
K_5	$\ln 2$	$\ln 2-\frac{1}{2}$	$\ln 2$	0

$$\frac{\mathbf{p}'^2+w^2}{8m_2^3}\left[\frac{2\pi i\delta(p_0')}{(p_0+i\epsilon)^2}-\frac{2\pi i\delta(p_0'-p_0)}{(p_0+i\epsilon)^2}-\frac{2\pi i\delta(p_0)}{(p_0'+i\epsilon)^2}\right]$$
$$+\frac{\mathbf{p}^2+w^2}{8m_2^3}\left[\frac{-2\pi i\delta(p_0')}{(p_0+i\epsilon)^2}-\frac{2\pi i\delta(p_0'-p_0)}{(p_0+i\epsilon)^2}\right.$$
$$\left.+\frac{2\pi i\delta(p_0)}{(p_0'+i\epsilon)^2}\right]+\frac{(\mathbf{p}-\mathbf{p}')^2+w^2}{8m_2^3}\left[-\frac{2\pi i\delta(p_0')}{(p_0+i\epsilon)^2}\right.$$
$$\left.+\frac{2\pi i\delta(p_0'-p_0)}{(p_0+i\epsilon)^2}-\frac{2\pi i\delta(p_0)}{(p_0'+i\epsilon)^2}\right]. \tag{46}$$

The explicit form of the three-photon interaction amplitudes of type (33) suggests that the terms in (46) yield corrections of the required order in α in the energy spectrum. Corrections of the same order in m_1/m_2 as in (46) originate in the quasipotential terms that contain the functions $f_{i1}(p,p')$ if for the muon denominators one uses the second approximate equality in (42). To draw some conclusion about the order of the corrections in the energy spectrum that are determined by these quasipotential terms, it is useful to subject the terms to certain transformations. For the sake of definiteness, in the functions $f_{i1}(p,p')$ we examine the massless terms proportional to p^2, p'^2, and pp':

$$3p^2\left[\frac{1}{D_\mu(p)}+\frac{1}{D_\mu(-p)}-\frac{1}{D_\mu(p-p')}-\frac{1}{D_\mu(p'-p)}\right]$$
$$+3p'^2\left[\frac{1}{D_\mu(-p')}+\frac{1}{D_\mu(p')}-\frac{1}{D_\mu(p'-p)}\right.$$
$$\left.-\frac{1}{D_\mu(p-p')}\right]-6pp'\left[\frac{1}{D_\mu(-p')}+\frac{1}{D_\mu(p')}\right.$$
$$\left.+\frac{1}{D_\mu(-p)}+\frac{1}{D_\mu(p)}-\frac{1}{D_\mu(p'-p)}-\frac{1}{D_\mu(p-p')}\right]$$
$$\approx\frac{3p^2}{2m_2}[-2\pi i\delta(p_0)+2\pi i\delta(p_0-p_0')]$$
$$+\frac{3p'^2}{2m_2}[-2\pi i\delta(p_0')+2\pi i\delta(p_0-p_0')]$$
$$-\frac{6pp'}{2m_2}[-2\pi i\delta(p_0)-2\pi i\delta(p_0')+2\pi i\delta(p_0-p_0')]. \tag{47}$$

The other terms in $f_{i1}(p,p')$ can be transformed in a similar manner. The next stage in the calculations amounts to integrating expressions like (46) or (47). A typical two-loop integral emerging in the process has the structure[17]

$$K_i=(4\pi)^2\int\frac{d^4p\,d^4p'}{-(2\pi)^8}\frac{G_i(p_0',p_0,m_1)P(\mathbf{p},\mathbf{p}',w)}{(p'^2-w^2+i\epsilon)[(p-p')^2+i\epsilon]}$$
$$\times\frac{1}{(p^2-w^2+i\epsilon)D_e(p')D_e(p)}, \tag{48}$$

where $G_i(p_0',p_0,m_1)$ necessarily contains a delta function, and $P(\mathbf{p}',\mathbf{p},w)$ a polynomial. In calculating (48), we used Feynman's parametrization to combine the denominators of the particle propagators, along with the symmetry of the integral under the interchange $p\leftrightarrow p'$. In this paper we have the following set of functions $G_i(p_0',p_0,m_1)$:

$$G_1=-\frac{2\pi i\delta(p_0-p_0')2m_1}{(p_0+i\epsilon)^2},\quad G_2=-\frac{2\pi i\delta(p_0)2m_1}{(p_0'+i\epsilon)^2},$$
$$G_3=-2\pi i\delta(p_0)2m_1,\quad G_4=-2\pi i\delta(p_0')2m_1,$$
$$G_5=-2\pi i\delta(p_0-p_0')2m_1. \tag{49}$$

The calculation of integrals like (48) with different G_i and P functions was done by Bodwin *et al.*[17] The results of calculations of the basis integrals (48) are listed in Table II.

Then the contributions determined by expressions (37)–(41) and (46) are, respectively,

$$\delta B_1^{3\gamma} = -\frac{1}{2}(Z\alpha)^6 \frac{m_1^2}{m_2}, \tag{50}$$

$$\delta B_2^{3\gamma} = (Z\alpha)^6 \frac{m_1^2}{m_2}\left(6\ln 2 - \frac{11}{48}\right). \tag{51}$$

The ''infrared'' logarithms ln$w$, which contain the photon mass (see the integrals K_i in Table II) introduced in (48) and emerge in the intermediate calculations, cancel out in the corrections $\delta B_1^{3\gamma}$ and $\delta B_2^{3\gamma}$.

We now examine the quasipotential terms containing the momenta of relative motion of the particles in the initial and final states, which we denote by $\mathbf{r}_1$ and $\mathbf{r}_2$. Allowing for these terms leads to the following corrections to the functions f_{i1}:

$$\Delta f_{21} = 10 m_1 p' r_2 + 5 p_0 p' r_2 + m_1 p r_2 + 3 p_0' p r_2, \tag{52}$$

$$\Delta f_{31} = -m_1 p' r_1 - 3 p_0 p' r_1 - 10 m_1 p r_1 - 5 p_0' p r_1, \tag{53}$$

$$\Delta f_{41} = m_1(7p' r_1 + 5p' r_2 + 10 p r_1 + 5 p r_2) + 6 p_0 p' r_1 + 5 p_0 p' r_2 + 8 p_0' p r_1 + 5 p_0' p r_2, \tag{54}$$

$$\Delta f_{51} = m_1(-5p' r_1 - 10 p' r_2 - 5 p r_1 - 7 p r_2) - 5 p_0 p' r_1 - 8 p_0 p' r_2 - 5 p_0' p r_1 - 6 p_0' p r_2, \tag{55}$$

$$\Delta f_{61} = m_1(-11p' r_1 - 11 p' r_2 - 11 p r_1 - 11 p r_2) - 8 p_0 p' r_1 - 8 p_0 p' r_2 - 8 p_0' p r_1 - 8 p_0' p r_2. \tag{56}$$

Allowing for the symmetry of the resulting integrals under simultaneous interchanges $p \leftrightarrow p'$ and $r_1 \leftrightarrow r_2$, we find that the integrals of the expressions (52)–(56) cancel. Hence the contribution of the relative motion of the particles to the fine structure in the required order in m_1/m_2 vanishes. Thus, the total value of the order $(Z\alpha)^6 m_1^2/m_2$ correction originating in the three-photon exchange diagrams for the S states of a hydrogen-like system is the sum of (50) and (51):

$$\Delta B_8 = \delta B_1^{3\gamma} + \delta B_2^{3\gamma} = (Z\alpha)^6 \frac{1}{n^3}\frac{\mu^3}{m_1 m_2}\left(6\ln 2 - \frac{35}{48}\right). \tag{57}$$

6. DISCUSSION

In this paper, using the diagrammatic quasipotential approach, we calculated all possible corrections of order $(Z\alpha)^6 m_1^2/m_2$ to the $n\,^3S_1$ levels of a hydrogen-like system; these were required for a comparison with the experimentally measured $2\,^3S_1 - 1\,^3S_1$ interval [see Eq. (2)]. Note that these corrections differ from the corrections of the corresponding order in the Lamb shift for the hydrogen atom.[13] Our final result, given by the sum of the ΔB_i terms in (4), (8), (11), (13), (24), (27), (32), and (57), is

$$\Delta B_{\rm tot} = \left(\frac{91}{12}\ln 2 - \frac{545}{144} - \frac{17}{6n} - \frac{37}{36n^2} + \frac{187}{96n^3}\right) \times \frac{m_1^2 (Z\alpha)^6}{m_2 n^3} + \varepsilon_n, \tag{58}$$

where

$$\varepsilon_n = \begin{cases} -\dfrac{23 m_1^2 (Z\alpha)^6}{12 m_2}, & n = 1, \\ -\dfrac{31 m_1^2 (Z\alpha)^6}{128 m_2}, & n = 2. \end{cases}$$

The numerical value of (58) for the ''large'' muonium fine-structure interval $2\,^3S_1 - 1\,^3S_1$ is 0.19 MHz. Earlier calculations of recoil corrections of order $(Z\alpha)^6 m_1^2/m_2^2$ for the S levels of the hydrogen atom were done in Refs. 6–9. The total contribution of the required order to the energy spectrum of the S states obtained by Eides and Grotch[8] via the Braun formula is

$$\Delta E_{\rm tot} = \left(\frac{1}{8} + \frac{3}{8n} - \frac{1}{n^2} + \frac{1}{2n^3}\right)\frac{(Z\alpha)^6 m_1^2}{m_2 n^3} + \left(4\ln 2 - \frac{7}{2}\right)\frac{(Z\alpha)^6 m_1^2}{m_2 n^3}. \tag{59}$$

The first term on the right-hand side, which contains a nontrivial dependence on the principal quantum number, was obtained by different authors taking different approaches.[6–9] In the approach based on Braun's formula, this contribution is singled out, and is governed by one-photon Coulomb exchange. In our approach, which uses the local quasipotential equation (3), corrections of this type originate in both the one-photon exchange quasipotential and the two-photon exchange quasipotential, as well as in second-order perturbation theory. Comparing (58) and (59), we see that the analytic expression (58) for the contribution differs somewhat from (59). The quasipotential of the S states of a hydrogen-like atom has the structure

$$V(r) = V_1(r) + (\boldsymbol{\sigma_1}\cdot\boldsymbol{\sigma}_2)V_2(r).$$

In our calculations we allowed for both the contribution of the first term in $V(r)$ to the energy spectrum and the second (spin-dependent) part of $V(r)$. Eides and Grotch[8] and Yelkhovsky[9] studied only the contribution of $V_1(r)$, so that the difference between (58) and (59) is perfectly understandable, and our results are consistent with those of Refs. 8 and 9. To a certain extent, it is useful to compare the numerical values of the corrections (58) and (59) for muonium levels with $n=1$ and $n=2$. These values are, respectively, -0.212 MHz and -0.065 MHz for $n=1$ and -0.021 MHz and -0.006 MHz for $n=2$. The contribution of the correction (58) to the hydrogen atom interval $2S-1S$ is 21.5 kHz, and the values of the corresponding contributions obtained in Refs. 8 and 9 are 6.6 kHz (Ref. 8) and 14.5 kHz (Ref. 9).

The authors are deeply grateful to A. S. Elkhovskiĭ and M. I. Éĭdes for critical remarks and interest in this work, to S. G. Karshenboĭm, É. A. Kuraev, P. Labelle, V. A. Saleev,

R. A. Sen'kov, and I. B. Khriplovich for stimulating discussions, to H. Grotch for sending a copy of Ref. 6, and to R. Mertig for sending the FeynCalc manual. This research was funded by the Russian Fund for Fundamental Research (Grant No. 98-02-16185) and the Universities of Russia Basic Research Program (Grant No. 2759).

[1)]The authors are grateful to S. N. Bagaev for discussing this problem.

*)E-mail: mart@info.ssu.samara.ru

[1] T. Kinoshita and J. Sapirstein, in *Proc. 9th Int. Conf. on Atomic Physics* (1984), p. 38.
[2] E. E. Salpeter, Phys. Rev. **87**, 328 (1952).
[3] H. Grotch and D. R. Yennie, Rev. Mod. Phys. **41**, 350 (1969).
[4] T. Fulton, Phys. Rev. A **26**, 1794 (1982).
[5] M. I. Eides and H. Grotch, Phys. Rev. A **52**, 1757 (1995).
[6] K. Pachucki and H. Grotch, Phys. Rev. A **51**, 1854 (1995).
[7] A. S. Elkhovskiĭ, Zh. Éksp. Teor. Fiz. **110**, 431 (1996) [JETP **83**, 230 (1996)].
[8] M. I. Eides and H. Grotch, Phys. Rev. A **55**, 3351 (1997).
[9] A. S. Yelkhovsky, Zh. Éksp. Teor. Fiz. **113**, 865 (1998) [JETP **86**, 472 (1998)].
[10] I. B. Khriplovich, A. I. Milstein, and A. S. Yelkhovsky, Phys. Lett. B **282**, 237 (1992).
[11] R. N. Fell, Phys. Rev. Lett. **68**, 25 (1992).
[12] R. N. Fell, I. B. Khriplovich, A. I. Milstein, and A. S. Yelkhovsky, Phys. Lett. A **181**, 172 (1993).
[13] V. V. Dvoeglazov, Yu. N. Tyukhtyaev, and R. N. Faustov, Fiz. Élem. Chastits At. Yadra **25**, 144 (1994) [Phys. Part. Nuclei **25**, 58 (1994)].
[14] S. Chu, A. P. Mills Jr., A. G. Yodh, K. Nagamine, Y. Miyake, and T. Kuga, Phys. Rev. Lett. **60**, 101 (1988).
[15] S. Chu, A. P. Mills Jr., and J. L. Hall, Phys. Rev. Lett. **52**, 1689 (1984).
[16] M. S. Fee, S. Chu, A. P. Mills Jr., R. J. Chichester, D. M. Zuckerman, E. D. Shaw, and K. Danzmann, Phys. Rev. A **48**, 192 (1993).
[17] G. T. Bodwin, D. R. Yennie, and M. A. Gregorio, Rev. Mod. Phys. **57**, 723 (1985).
[18] A. I. Mil'shteĭn and I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **106**, 689 (1994) [JETP **79**, 379 (1994)].
[19] W. E. Caswell and G. P. Lepage, Phys. Rev. A **18**, 810 (1978).
[20] V. A. Rizov and I. T. Todorov, Fiz. Élem. Chastits At. Yadra **6**, 669 (1975) [Sov. J. Part. Nucl. **6**, 269 (1975)].
[21] A. P. Martynenko and R. N. Faustov, Yad. Fiz. **45**, 770 (1987) [Sov. J. Part. Nucl. **45**, 479 (1987)].
[22] R. N. Faustov and A. P. Martynenko, Yad. Fiz. **60**, 1407 (1997) [Phys. At. Nucl. **60**, 1272 (1997)].
[23] A. P. Martynenko and R. N. Faustov, Teoret. Mat. Fiz. **66**, 399 (1986) [Teor. Mat. Fiz. **66**, 264 (1986)].
[24] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed., Pergamon Press, Oxford (1991).
[25] A. Messiah, *Quantum Mechanics* [in Russian], Vol. 2, Nauka, Moscow (1979) [French orig.: *Méchanique Quantique*, Dunod, Paris (1959); English transl.: North-Holland, Amsterdam (1962)].
[26] J. Schwinger, J. Math. Phys. **5**, 1606 (1964).
[27] W. Buchmuller and E. Remiddi, Nuovo Cimento A **60**, 109 (1980).
[28] R. Hostler, J. Math. Phys. **5**, 591 (1964).
[29] M. B. Voloshin, Yad. Fiz. **36**, 247 (1982) [Sov. J. Nucl. Phys. **36**, 143 (1982)].
[30] S. G. Karshenboĭm, Zh. Éksp. Teor. Fiz. **103**, 1105 (1993) [JETP **76**, 541 (1993)].
[31] S. A. Zapryagaev, N. L. Manakov, and V. G. Pal'chikov, *Theory of One- and Two-Electron Multiply-Charged Ions* [in Russian], Énergoatomizdat, Moscow (1985).
[32] V. G. Ivanov and S. G. Karshenboĭm, Zh. Éksp. Teor. Fiz. **109**, 1219 (1996) [JETP **82**, 656 (1996)].
[33] S. Wolfram, *Mathematica—A System for Doing Mathematics by Computer*, Addison-Wesley, Reading, MA (1988).

Translated by Eugene Yankovsky

Stabilization of an atom in a strong laser field

A. Ya. Kazakov*)

St. Petersburg University of Aerospace Instrumentation, 190000 St. Petersburg, Russia
(Submitted 30 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1236–1242 (April 1999)

A quasiresonant laser field initiates the decay of an initially occupied atomic level into the continuum. If the amplitude of the external field is sufficiently high, other atomic levels, not meeting the condition for exact resonance, begin to participate in the atomic dynamics. This phenomenon leads to the stabilization of the atom. © *1999 American Institute of Physics.* [S1063-7761(99)00604-6]

1. INTRODUCTION

During recent years extensive research has been conducted in the field of the dynamics of an atom in intense and superintense laser light. A nontrivial effect in this field of quantum physics is the stabilization of the atom. Suppose that an atom is placed in a laser field whose frequency satisfies the resonance condition, i.e., the field is in resonance with the transition from the occupied atomic level to the continuum. From a ''naive'' standpoint, the greater the amplitude of the external field, the faster the initially occupied atomic level will decay into the continuum. The analysis of a decay of an atom into the continuum initiated by a resonant laser leads to the relationship $\kappa \propto r^2$, where κ is the corresponding relaxation constant, and r is the amplitude of the external field in appropriate units (see, e.g., Ref. 1). However, a careful study shows that the dependence of the relaxation constant on the external-field amplitude may be extremely complex.[2–7] Various physical phenomena leading to a nontrivial behavior of $\kappa(r)$, such as resonant stabilization, adiabatic stabilization, and interference stabilization, have been described in the literature (a discussion of this aspect and the literature can be found in Ref. 2). This paper discusses one more physical phenomenon that effectively leads to stabilization of the atom.

Let us take an initially occupied atomic level. An external laser field initiates the decay of the level into the continuum. We assume that there is also a set of atomic levels out of exact resonance with the specified level and the continuum. However, the very notion of exact resonance originates in perturbation theory. As the amplitude of the external field increases, the levels (from the perturbation-theory viewpoint) that were not in exact resonance with the occupied level begin to effectively interact with that level. We will see that this interaction can significantly alter the ionization of the atom.

2. DYNAMICS OF THE PHYSICAL SYSTEM

We study the simplest situation by assuming that there is only one ''additional'' level, which can begin to interact with the initially occupied level as the amplitude of the external field increases (see the level diagram in Fig. 1). The dynamics of such a system is described by the Schrödinger equation

$$i\hbar \frac{\partial \Psi(t)}{\partial t} = [H_0 + xF(t)]\Psi(t), \tag{1}$$

with the external field $F(t) = R\cos \Omega t$. Let $|0\rangle$, $|1\rangle$, and $|E\rangle$ be the wave functions of the initially occupied level, the additional level, and the states in the continuum, $E \in K = [E_c, \infty]$. We expand the wave function of the atom in this set of states:

$$\Psi(t) = A_0(t)|0\rangle + A_1(t)|1\rangle + \int_K B(E,t)|E\rangle \, dE. \tag{2}$$

If we substitute this expansion into Eq. (1), we obtain a set of equations

$$i\hbar A_0'(t) = E_0 A_0(t) + R\cos(\Omega t) \times \left\{ g_1 A_1(t) + \int_K B(E,t) g(E)\, dE \right\},$$

$$i\hbar A_1'(t) = E_1 A_1(t) + R\cos(\Omega t)\, g_1 A_0(t),$$

$$i\hbar B'(E,t) = EB(E,t) + R\cos(\Omega t)\, g(E) A_0(t),$$

$$C'(t) \equiv \frac{\partial C(t)}{\partial t},$$

where g_1 and $g(E)$ are the corresponding elements of the atomic dipole-moment operator. We assume that the frequency Ω of the external field (the field that transfers state $|0\rangle$ into the continuum) is much higher than all other frequency parameters. This assumption makes it possible to employ the rotating wave approximation (RWA). Using the substitutions

$$B(E,t) = \exp\{-i(\Omega + E_0/\hbar)t\} b(E,t),$$

$$A_0(t) = \exp\{-iE_0 t/\hbar\}\, a_0(t),$$

$$A_1(t) = \exp\{-i(E_0/\hbar - \Omega)t\}\, a_1(t)$$

and introducing the variable ς via $E = (\varsigma + \Omega)\hbar + E_0$, we can eliminate the optical frequency:

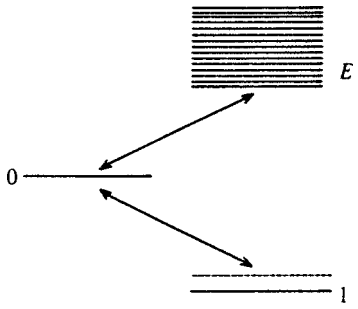

FIG. 1.

$$ia_0'(t)=r\left[g_1a_1(t)+\int_K g(\zeta)b(\zeta,t)\,d\zeta\right], \tag{3}$$

$$ia_1'(t)=\Delta_1a_1(t)+rg_1a_0(t), \tag{4}$$

$$ib'(\varsigma,t)=\zeta b(\zeta t)+rg(\zeta)a_0(t), \tag{5}$$

where $r=R/2\hbar$, and $\Delta_1=(E_1-E_0)/\hbar+\Omega$ is the offset of the laser-light frequency from the frequency of the transition between the discrete levels. (For objects related to ς we use the same notation as for objects related to the variable E.) Here we are interested in the solution of the system of equations (3)–(5) with the initial conditions

$$a_0(0)=1, \tag{6}$$

$$a_1(0)=0, \tag{7}$$

$$b(\varsigma,0)=0. \tag{8}$$

Using the initial condition (8) and integrating Eq. (5), we find that

$$b(\varsigma,t)=-irg(\zeta)\int_0^t\exp[i\zeta(x-t)]\,a_0(x)\,dx. \tag{9}$$

Substituting this into Eq. (3) yields

$$a_0'(t)=-irg_1a_1(t)-r^2\int_0^t a_0(x)Q(t-x)\,dx, \tag{10}$$

$$Q(y)=\int_K g^2(\zeta)\exp[-iy\zeta]\,d\zeta.$$

Equation (10), Eq. (4), and the initial conditions (6) and (7) constitute a complete system of equations for finding the functions $a_0(t)$ and $a_1(t)$. This system of equations can by solved by the Laplace transform method. However, we will use a different approach, applicable also in the case of nontrivial modulation of the external field. More precisely, we resort to certain asymptotic considerations.

Let us examine the structure of the function $Q(y)$. We assume that $g(\zeta)$ changes significantly only under variations of its argument (we denote the corresponding parameter by D) that are much larger than other energy parameters (or, in corresponding units of measurement, frequency parameters) of the problem. Thus, $g(\zeta)=v(\zeta/D)$, with $v'(y)$ and $v''(y)$ being of the same order as $v(y)$ for $y=O(1)$. Let $\varsigma=Ds$ and $Q(y)=Dq(y)$, where

$$q(y)=\int_{K_1}v^2(s)\exp[-iyDs]\,ds.$$

Under our assumptions, $q(y)$ is the ''fast'' (or rapidly varying) function. The second term on the right-hand side of Eq. (10) is the integral of the product of the ''fast'' and ''slow'' functions. To obtain an asymptotic expansion [in the small parameter $\max(|rg_1|,|\Delta_1|)/D$)] of such an integral, it suffices to integrate by parts.[8] After the first step has been completed, instead of (10) we have

$$a_0'(t)=-irg_1a_1(t)-r^2a_0(t)[iS+\pi g^2(0)], \tag{11}$$

where

$$S=\text{V.P.}\int_{K_1}\frac{v^2(s)}{s}ds$$

is the corresponding Stark shift of the edge of the continuum.

Note that we derived Eq. (11) by assuming that $\Omega\gg D\gg\max(|rg_1|,|\Delta_1|)$. This means that we set up the leading term in the expansion of the solution of the initial problem in the small parameters D/Ω and $(rg_1,|\Delta_1|)/D$. But Eq. (11) is valid even if $D\gg\Omega\gg rg_1,\ |\Delta_1|$. In this case to derive the equation we must reverse the order of operations: we first expand the analog of Eq. (10) in the small parameter Ω/D, and then, employing RWA (i.e., eliminating the optical parameter), set up the leading term in the expansion of the solution in the small parameter $(rg_1,|\Delta_1|)/\Omega$. Here the possibility of using RWA can be justified by standard means. We leave out the mathematical details, since it is are similar to the approach discussed in Ref. 8.

We write the system of equations (4) and (11) in the form

$$A'(t)=\Lambda A(t), \tag{12}$$

$$A=\begin{pmatrix}a_0(t)\\a_1(t)\end{pmatrix},\quad \Lambda=\begin{pmatrix}-r^2[iS+\pi g^2(0)] & -irg_1\\ -irg_1 & -i\Delta_1\end{pmatrix},$$

$$A(0)=\begin{pmatrix}1\\0\end{pmatrix}. \tag{13}$$

The eigenvalues of the matrix Λ can easily be calculated:

$$\lambda_1=-\pi r^2g^2(0)/2-i[\Delta_1+r^2S]/2 +[(\pi r^2g^2(0)+i(r^2S-\Delta_1))^2-r^2g_1^2]^{1/2}, \tag{14}$$

$$\lambda_2=-\pi r^2g^2(0)/2-i[\Delta_1+r^2S]/2 -[(\pi r^2g^2(0)+i(r^2S-\Delta_1))^2-r^2g_1^2]^{1/2}. \tag{15}$$

Let ψ_1 and ψ_2 be the corresponding eigenvectors (we will not write the expressions for these eigenvectors, which are states of the atom ''dressed'' by the field). We wish to note, however, that $\psi_2=A(0)$ at $r=0$.

Figures 2 and 3 depict typical plots of the functions $\kappa_1(r)=-\text{Re}\lambda_1(r)$ and $\kappa_2(r)=-\text{Re}\lambda_2(r)$ of the corresponding relaxation constants. Figure 4 depicts the ratios κ_2/κ_1 as functions of r. As noted in Ref. 9, the real values of the parameters $g(0)$ and g_1 are poorly known. For this reason we used estimates: $g(0)=0.2$, $g_1=1.5$, $S=1$, $\Delta_1=-8$

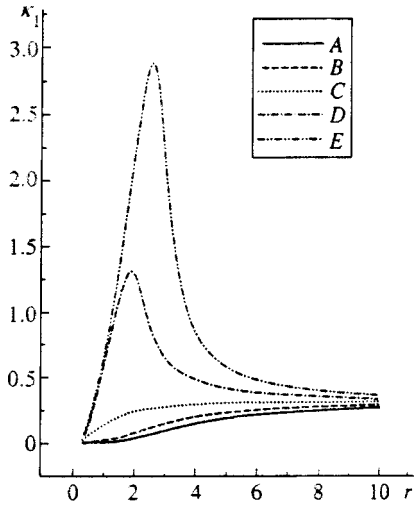

FIG. 2.

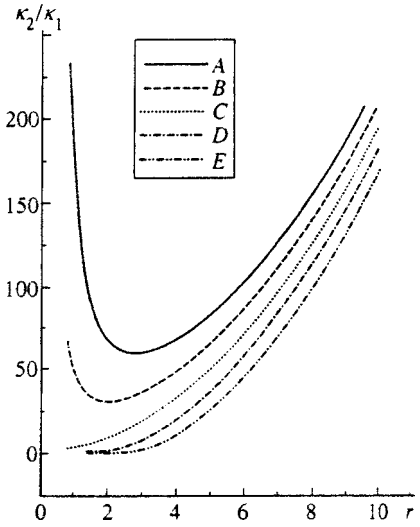

FIG. 4.

(for curves *A*), $\Delta_1=-4$ (for curves *B*), $\Delta_1=0$ (for curves *C*), $\Delta_1=4$ (for curves *D*), and $\Delta_1=8$ (for curves *E*). Note that the curves representing $\kappa_2(r)$ for our values of Δ_1 are essentially identical. These diagrams suggest that as the amplitude r of the external field increases, the values of the relaxation constants begin to differ appreciably. When the effective Rabi parameter rg_1 becomes of order Δ_1, which is the offset of the additional level from resonance, the relaxation constants differ already by an order of magnitude. Within a broad range of the parameters, the ratio of these quantities assumes values of order $10–10^2$or 10^{-1} (see Fig. 4). Although the choice of the values of the parameters $g(0)$, g_1, and S is random (to a certain extent) and hence the problem is actually a model, similar behavior of these curves can be observed within a broad range of values of $g(0)$, g_1, and S. Note that disparate physical processes can be examined by this model, such as the ionization of Rydberg atoms, photoejection of an electron from an ion, and ionization of ''atomlike'' quantum well systems. Of course, the parameters of the model in these cases may be quite different.

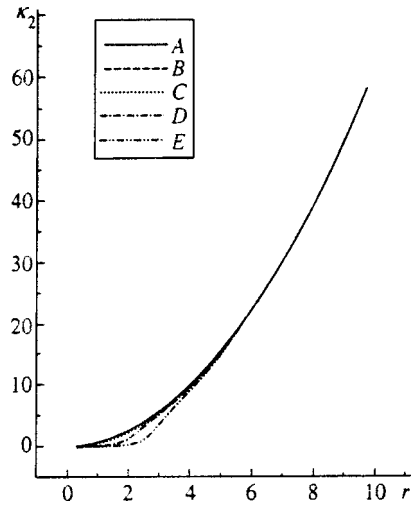

FIG. 3.

In light of these results, let us discuss the dependence of the ionization of our model atom on the amplitude of the external field. The solution of the initial problem specified by (12) and (13) can be written

$$A(t)=\exp(\lambda_1 t)\,p_1\psi_1+\exp(\lambda_2 t)\,p_2\psi_2, \tag{16}$$

where p_1 and p_2 are the coefficients of the expansion of $A(0)$ in the basis ψ_1,ψ_2. As noted earlier, for small amplitudes of the external field, p_2 is close to unity and p_1 is close to zero. However, p_1 increases with r, and when rg_1 becomes of order Δ_1, p_1 becomes comparable to unity. Then, as Eq. (16) implies, the rate of decay of the population of the atom into the continuum is determined by both relaxation constants. Here the part of the population referring to the ''dressed'' state ψ_1 decays much slower than the part corresponding to the ''dressed'' state ψ_2. This phenomenon can be interpreted as stabilization of the atom in a strong laser field.

3. CONCLUSION

We have discussed one more physical mechanism for the stabilization of an atom in a strong laser field that differs from those discussed earlier (see, e.g., Ref. 2). We have found that the existence of a group of levels ''close'' to resonance with an initially occupied level can lead to trapping of the population of the atom and substantial alteration of the ionization process. Formally this manifests itself in the system being characterized by several relaxation constants that differ substantially in order of magnitude. (Note that the largest relaxation constant for such a system is close to the relaxation constant for the level–continuum system calculated by ordinary perturbation theory, $\kappa=\pi r^2 g^2(0)$; see Ref. 1). Of course, the case we have examined is only a model, but a more general physical system exhibits similar properties (e.g., a system with several ''additional'' levels or a system in which some of these levels interact with the continuum). In real situations the stabilization of an atom may occur for various reasons or even a group of reasons. Note that our range of parameters ($rg_1 \ll D \ll \Omega$) differs from that

discussed in Ref. 6 ($D \ll r g_1 \ll \Omega$) and Ref. 7 ($D, \Omega \ll r_{\text{eff}}$, where r_{eff} is a parameter interpreted as the effective Rabi parameter in the case of ultrastrong fields).

After this paper had been prepared for publication, there appeared Poluéktov and Fedorov's paper[10] in which ''interference'' stabilization was discussed for Λ and V systems. Formally, the problem discussed in Ref. 10 (and in the earlier paper in Ref. 9) is close to the one discussed in the present paper. Both are based on the same fact: a strong laser field induces states of the atom that are ''dressed'' by the field (we denote these states by Ψ_f) and have a small decay rate. In ''interference'' stabilization, such states emerge because of the interference of transitions from different atomic levels into the continuum and play an important role at all values of the external-field amplitudes. In our case, when the amplitude of the external field is low, these states have almost no effect on the dynamics of the atom; only when the amplitude is large do they begin to determine the dynamics. There are two reasons for this. First, in our case the very structure of the Ψ_f-states (the state is a linear combination of the unperturbed states of the atom) depends more strongly on the amplitude of the external field than it does in the case of ''interference'' stabilization. Second, within the proposed stabilization mechanism, the coefficient p_2 in (16) (which describes the relationship between the initial state of the atom and Ψ_f) strongly depends on the amplitude of the external field (which tends to zero as r decreases), while its analog in the case of ''interference'' stabilization does not undergo such substantial changes.

*)E-mail: akaz@phsc2.stu.neva.ru

[1] V. M. Akulin and N. V. Karlov, *Intense Resonant Interactions in Quantum Electrodynamics*, Springer-Verlag, Berlin (1991).
[2] *Super-Intense Laser–Atom Physics, Proc. NATO Adv. Res. Workshop IV*, H. G. Muller and M. V. Fedorov (eds.), Kluwer Academic Publishers, Dordrecht (1996).
[3] M. V. Fedorov, *Atomic and Free Electrons in a Strong Light Field*, World Scientific, River Edge, N.J. (1997).
[4] K. Burnett, V. C. Reed, and P. L. Knight, J. Phys. B **26**, 561 (1993).
[5] N. B. Delone and V. P. Krainov, *Multiphoton Processes in Atoms*, Springer-Verlag, New York (1994).
[6] A. Ya. Kazakov, Zh. Éksp. Teor. Fiz. **107**, 1047 (1995) [JETP **80**, 586 (1995)].
[7] A. Ya. Kazakov, J. Phys. A **31**, 3019 (1998).
[8] M. V. Fedoryuk, *Asymptotic Expressions, Integrals and Series* [in Russian], Nauka, Moscow (1987).
[9] M. V. Fedorov, ''Interference stabilization,'' in *Super-Intense Laser–Atom Physics, Proc. NATO Adv. Res. Workshop IV*, H. G. Muller and M. V. Fedorov (eds.), Kluwer Academic Publishers, Dordrecht (1996), p. 11.
[10] N. P. Poluéktov and M. V. Fedorov, Zh. Éksp. Teor. Fiz. **114**, 821 (1998) [JETP **87**, 445 (1998)].

Translated by Eugene Yankovsky

Pseudogap phase formation in the crossover from Bose–Einstein condensation to BCS superconductivity

V. P. Gusynin*) and V. M. Loktev†)

Bogolyubov Institute for Theoretical Physics, 252143 Kiev, Ukraine

S. G. Sharapov

Department of Physics, University of Pretoria, 0002 Pretoria, South Africa

(Submitted 9 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 1243–1262 (April 1999)

A phase diagram for a 2D metal with variable carrier density has been derived. It consists of a normal phase, where the order parameter is absent: a so-called ''abnormal normal'' phase where this parameter is also absent but the mean number of composite bosons (bound pairs) exceeds the mean number of free fermions; a pseudogap phase where the absolute value of the order parameter gradually increases but its phase is a random value, and finally a superconducting (here Berezinskiĭ–Kosterlitz–Thouless) phase. The characteristic transition temperatures between these phases are found. The chemical potential and paramagnetic susceptibility behavior as functions of the fermion density and the temperature are also studied. An attempt is made to qualitatively compare the resulting phase diagram with the features of underdoped high-T_c superconducting compounds above their critical temperature.
 [S1063-7761(99)00704-0]

1. INTRODUCTION

The study of the crossover region between superconductivity of Cooper pairs and superfluidity of composite bosons is attracting much attention due to its close relationship to the problem of describing high-temperature superconductors (HTSC) (see, e.g., Refs. 1–3). At present this region is understood for 3D systems, both at zero and finite temperatures.[4,5] The crossover in quasi-2D systems has also been studied,[6] albeit only partially, whereas for 2D systems only the case of $T=0$ has been studied thoroughly.[4,7] This is related to the fact that fluctuations of the charged complex order parameter in 2D systems are so large that they destroy long-range order at any finite temperature (Coleman–Mermin–Wagner–Hohenberg (CMWH) theorem[8]). In this case the appearance of an inhomogeneous condensate with a power-law decay for the correlations (the so-called Berezinskiĭ–Kosterlitz–Thouless (BKT) phase) is possible. However an adequate mathematical description for BKT phase formation is still lacking.

Most previous analyses[9–11] of the behavior of 2D systems at $T\neq 0$ have been based on the Nozières–Schmitt–Rink approach.[12] This approach is simply a Gaussian approximation to the functional integral, and this perhaps explains the difficulties faced in these calculations. On the one hand, Gaussian fluctuations destroy long-range order in 2D and if one searches for the T_c^{2D} at which the order sets in, one should obtain zero in accordance with the aforementioned theorems.[8] On the other hand, taking Gaussian fluctuations into account is completely inadequate to describe the BKT transition.[13]

Nonetheless, there has been some progress. For example, the BKT transition has been studied in relativistic $2+1$-theory,[14] and the crossover from superconductivity to superfluidity has been considered[15] as a function of the carrier density n_f (see also Ref. 16). However, the method employed in Ref. 15 to obtain the temperature $T_{\rm BKT}$ has several drawbacks. Most importantly, the equation for $T_{\rm BKT}$ was obtained without considering the existence of a neutral (real) order parameter ρ, whose appearance at finite T does not violate the CMWH theorem.

As we show below, ρ defines the modulus of a multivalued complex order parameter Φ for a 2D system. As a result of allowing for a neutral order parameter, a region where ρ decays gradually to zero appears in the phase diagram of the system. This region separates the standard normal phase with $\rho=0$ from the BKT phase, where the correlations exhibit power-law decay. Despite the exponential decay of correlations, this new region of states may be expected to possess unusual properties, since ρ plays the same role as the energy gap Δ in the theory of ordinary superconductors in many cases.[1)] The possible existence of such a phase, which in some sense is also normal, may shed light on the anomalous behavior of the normal state of HTSC (see, for example, the reviews in Refs. 1, 2, and 18). In particular, the temperature dependencies of the spin susceptibility, resistivity, specific heat, photoemission spectra, and other quantities[2,19] can be explained by the formation of either a pseudogap or a spin gap in the region $T>T_c$.

Using a very simple continuum 2D model, this approach was first attempted in a brief note,[20] where we calculated $T_{\rm BKT}$ and T_ρ (T_ρ is the temperature defined by the condition $\rho=0$) self-consistently as functions of n_f, and established the boundaries of this new *pseudogap* region, which lies between $T_{\rm BKT}$ and T_ρ.

The purpose of this article is to develop this approach further. Using the static paramagnetic susceptibility as an example, we demonstrate that the pseudogap opens below T_ρ. Furthermore, we analyze the difference between the commonly used (see Refs. 3 and 4) pairing temperature T_p and the temperature T_ρ introduced here. These temperatures turn out to be different if the chemical potential $\mu<0$. We also introduce here an *abnormal normal* phase, which lies between T_p and T_ρ, where performed bosons exist. This more detailed study helps to clarify the physical import of T_ρ, as well as the nature of the transition at T_ρ. It was believed in the related model[14] that this is a second-order phase transition. We argue however, that fluctuations in the phase of the order parameter can transform the transition to a crossover, as observed experimentally.

In Sec. 2 we present the model and the relevant formalism. The equations for $T_{\rm BKT}$, ρ, T_ρ, and the chemical potentials $\mu(T_{\rm BKT})$ and $\mu(T_\rho)$ are derived in Sec. 3. Since the technique employed to obtain the equation for $T_{\rm BKT}$ is not widely used, we consider it useful to present a detailed derivation of this equation. (The details of the calculation of the effective potential and useful series are given in Appendix A.) The systems of equations for $T_{\rm BKT}$, $\rho(T_{\rm BKT})$, $\mu(T_{\rm BKT})$ and T_ρ, $\mu(T_\rho)$ are analyzed in Sec. 4. The difference between pairing temperature T_p and the temperature T_ρ is discussed in Sec. 5. Also discussed is the physical import of T_ρ. Using the example of the static spin susceptibility, it is shown in Sec. 6 that the resulting pseudogap phase can in fact be used to explain the aforementioned anomalous properties of HTSC.

2. THEORETICAL FRAMEWORK

The simplest model Hamiltonian density for fermions confined to a 2D volume v is[4,7,9]

$$\mathscr{H}=\psi^\dagger_\sigma(x)\left(-\frac{\nabla^2}{2m}-\mu\right)\psi_\sigma(x)-V\psi^\dagger_\uparrow(x)\psi^\dagger_\downarrow(x)\psi_\downarrow(x)\psi_\uparrow(x), \tag{2.1}$$

where $x\equiv\mathbf{r},\tau;\psi_\sigma(x)$ is a fermion field, m is the effective fermion mass, μ is the chemical potential, and V is an effective local attraction constant; we take $\hbar=k_B=1$.

The Hubbard–Stratonovich method, which is standard for these problems,[21] can be applied to write the partition function $Z(v,\mu,T)$ as a functional integral over Fermi fields (Nambu spinors) and the auxiliary field $\mathbf{\Phi}=V\psi^\dagger_\uparrow\psi^\dagger_\downarrow$. In contrast to the usual method for calculating Z in $\mathbf{\Phi}$, $\mathbf{\Phi}$* variables, the parametrization $\Phi(x)=\rho(x)\exp[-i\theta(x)]$ is more appropriate for presenting the corresponding integral in two dimensions[22] (see also Refs. 23 and 24). When this replacement by modulus-phase variables is implemented, it is evident that one must also replace $\psi_\sigma(x)=\chi_\sigma(x)\exp[i\theta(x)/2]$. Physically, this amounts to replacing the charged fermion $\psi_\sigma(x)$ with a neutral fermion $\chi_\sigma(x)$ and spinless charged boson $e^{i\theta(x)/2}$. Note that while one may formally use any self-consistent definition of the new variables, the physical condition that the macroscopic variable $\mathbf{\Phi}(x)$ be single-valued under 2π rotations fixes the parametrization. This was not taken into account in Ref. 20, where a different parametrization was used.

As a result, one obtains

$$Z(v,\mu,T)=\int\rho\mathscr{D}\rho\mathscr{D}\theta\exp[-\beta\Omega(v,\mu,T,\rho(x),\partial\theta(x))], \tag{2.2}$$

where

$$\beta\Omega(v,\mu,T,\rho(x),\partial\theta(x))=\frac{1}{V}\int_0^\beta d\tau\int d\mathbf{r}\rho^2(x)-\mathrm{Tr}\ln G^{-1}+\mathrm{Tr}\ln G_0^{-1} \tag{2.3}$$

is the one-loop effective action, which depends on the modulus-phase variables. The action (2.3) can be expressed in terms of the Green function of the initial (charged) fermions, which in the new variables has the operator form

$$G^{-1}=-\hat{I}\partial_\tau+\tau_3\left(\frac{\nabla^2}{2m}+\mu\right)+\tau_1\rho(\tau,\mathbf{r})-\tau_3\left[\frac{i\partial_\tau\theta(\tau,\mathbf{r})}{2}+\frac{(\nabla\theta(\tau,\mathbf{r}))^2}{8m}\right]+\hat{I}\left[\frac{i\nabla^2\theta(\tau,\mathbf{r})}{4m}+\frac{i\nabla\theta(\tau,\mathbf{r})\nabla}{2m}\right]. \tag{2.4}$$

The free fermion Green function $G_0=G|_{\mu,\rho,\theta=0}$ provides a convenient regularization in the process of calculation. It is important that neither the smallness nor slowness of the variation of the phase of the order parameter is assumed in obtaining expression (2.3). In other words, it is formally exact.

Since the low-energy dynamics of phases for which $\rho\neq0$ is governed mainly by long-wavelength fluctuations of $\theta(x)$, only the lowest-order derivatives of the phase need be retained in the expansion of $\Omega(v,\mu,T,\rho(x),\partial\theta(x))$:

$$\Omega(v,\mu,\rho(x),\partial\theta(x))\simeq\Omega_{\rm kin}(v,\mu,T,\rho,\partial\theta(x))+\Omega_{\rm pot}(v,\mu,T,\rho), \tag{2.5}$$

where

$$\Omega_{\rm kin}(v,\mu,T,\rho,\partial\theta(x))=T\ \mathrm{Tr}\sum_{n=1}^{\infty}\frac{1}{n}(\mathscr{G}\Sigma)^n\Big|_{\rho=\rm const} \tag{2.6}$$

and

$$\Omega_{\rm pot}(v,\mu,T,\rho)=\left(\frac{1}{V}\int d\mathbf{r}\rho^2-T\ \mathrm{Tr}\ \ln\ \mathscr{G}^{-1}+T\ \mathrm{Tr}\ln G_0^{-1}\right)\Big|_{\rho=\rm const}. \tag{2.7}$$

The kinetic $\Omega_{\rm kin}$ and potential $\Omega_{\rm pot}$ parts can be expressed in terms of the Green function of the neutral fermions, which satisfies the equation

$$\left[-\hat{I}\partial_\tau+\tau_3\left(\frac{\nabla^2}{2m}+\mu\right)+\tau_1\rho\right]\mathscr{G}(\tau,\mathbf{r})=\delta(\tau)\delta(\mathbf{r}), \tag{2.8}$$

and the operator

$$\Sigma(\partial\theta)\equiv\tau_3\left[\frac{i\partial_\tau\theta}{2}+\frac{(\nabla\theta)^2}{8m}\right]-\hat{I}\left[\frac{i\nabla^2\theta}{4m}+\frac{i\nabla\theta(\tau,\mathbf{r})\nabla}{2m}\right]. \tag{2.9}$$

The representation (2.5) enables one to obtain the full set of equations necessary to find $T_{\rm BKT}$, $\rho(T_{\rm BKT})$, and $\mu(T_{\rm BKT})$ at given ϵ_F (or, for example, $\rho(T)$ and $\mu(T)$ at given T and ϵ_F). While the equation for $T_{\rm BKT}$ will be written using the kinetic part (2.6) of the effective action, the equations for $\rho(T_{\rm BKT})$ and $\mu(T_{\rm BKT})$ [or $\rho(T)$ and $\mu(T)$] can be obtained using the mean field potential (2.7): It turns out that at a phase for which $\rho\neq 0$, the mean-field approximation for the modulus variable describes the system quite well. This is mainly related to the nonperturbative character of the Hubbard–Stratonovich method, i.e., most effects carry over for a nonzero value of ρ.

It is clear that the CMWH theorem does not preclude nonzero $\langle\rho\rangle$ and, as a consequence, an energy gap for fermion χ, since no continuous symmetry is broken when such a gap appears. Despite strong phase fluctuations in the two-dimensional case, the energy gap in the spectrum of the neutral fermion χ can still persist in the spectrum of the charged fermion ψ,[22] even well above the critical temperature.[2)] We believe that the pseudogap widely discussed in high-T_c cuprates might be attributable to the energy gap of a neutral fermion introduced in the way described above, so that the pseudogap itself can be considered a remnant of the superconducting gap. The condensate of neutral fermions has nothing to do with the superconducting transition; the latter is only possible when the superfluid density of bosons becomes large enough to stiffen the phase $\theta(x)$. The temperature T_ρ at which nonzero $\langle\rho\rangle$ develops should be identified in this approach with the pseudogap onset temperature.[2,19] The strategy of treating charge and spin degrees of freedom as independent seems to be quite useful, and at the same time a very general feature of two-dimensional systems.

3. DERIVATION OF SELF-CONSISTENT EQUATIONS FOR $T_{\rm BKT}$, NEUTRAL ORDER PARAMETER, AND CHEMICAL POTENTIAL

If the model under consideration is reduced to some known model describing the BKT phase transition, one can easily write the equation for $T_{\rm BKT}$, which in the present approach can be identified with the superconducting transition temperature T_c. Indeed, in the lowest orders the kinetic term (2.6) coincides with the classical spin XY-model,[25,26] which has the continuum Hamiltonian

$$\mathscr{H}=\frac{J}{2}\int d\mathbf{r}[\nabla\theta(\mathbf{r})]^2. \tag{3.1}$$

Here J is some coefficient (in the original classical discrete XY-model it is the stiffness of the relatively small spin rotations) and θ is the angle (phase) of the two-component vector in the plane.

The temperature of the BKT transition is, in fact, known for this model:

$$T_{\rm BKT}=\frac{\pi}{2}J. \tag{3.2}$$

Despite the very simple form[3)] of Eq. (3.2), it was derived (see, e.g., Refs. 25 and 26) using the renormalization group technique, which takes into account the non-single-valuedness of the phase θ. Thus, fluctuations of the phase are taken into account in a higher approximation than Gaussian. The XY-model was assumed to be adequate for a qualitative description of the underdoped cuprates[27] (see also Ref. 28), and the relevance of the BKT transition to Bose- and BCS-like superconductors was recently discussed in Ref. 16.

To expand $\Omega_{\rm kin}$ up to $\sim(\nabla\theta)^2$, it is sufficient to restrict ourselves to terms with $n=1,2$ in the expansion (2.6). The calculation is similar to that employed in Ref. 29, where only high densities n_f were considered at $T=0$. Thus, to obtain the kinetic part, one should directly calculate the first two terms of the series (2.6), which can be formally written $\Omega^{(1)}_{\rm kin}=T{\rm Tr}(\mathscr{G}\Sigma)$ and $\Omega^{(2)}_{\rm kin}=(1/2)T{\rm Tr}(\mathscr{G}\Sigma\mathscr{G}\Sigma)$. We note that Σ has the structure $\Sigma=\tau_3 O_1+\hat{I}O_2$, where O_1 and O_2 are differential operators [see (2.9)]. One can see, however, that the part of Σ proportional to the unit matrix $\hat{I}$ does not contribute to $\Omega^{(1)}_{\rm kin}$. Hence,

$$\Omega^{(1)}_{\rm kin}=T\int_0^\beta d\tau\int d\mathbf{r}\frac{T}{(2\pi)^2}\sum_{n=-\infty}^{\infty}\int d\mathbf{k}\ \ {\rm Tr}[\mathscr{G}(i\omega_n,\mathbf{k})\tau_3]\times\left(\frac{i\partial_\tau\theta}{2}+\frac{(\nabla\theta)^2}{8m}\right), \tag{3.3}$$

where

$$\mathscr{G}(i\omega_n,\mathbf{k})=-\frac{i\omega_n\hat{I}+\tau_3\xi(\mathbf{k})-\tau_1\rho}{\omega_n^2+\xi^2(\mathbf{k})+\rho^2} \tag{3.4}$$

is the Green function of neutral fermions in the frequency-momentum representation, with $\xi(\mathbf{k})=\varepsilon(\mathbf{k})-\mu$ and $\varepsilon(\mathbf{k})=\mathbf{k}^2/2m$.

The summation over the Matsubara frequencies $\omega_n=\pi(2n+1)T$ and integration over $\mathbf{k}$ in (3.3) can be easily performed using the sum (A7); thus one obtains

$$\Omega^{(1)}_{\rm kin}=T\int_0^\beta d\tau\int d\mathbf{r}n_F(\mu,T,\rho)\left(\frac{i\partial_\tau\theta}{2}+\frac{(\nabla\theta)^2}{8m}\right), \tag{3.5}$$

where

$$n_F[\mu,T,\rho(\mu,T)]=\frac{m}{2\pi}\left\{\sqrt{\mu^2+\rho^2}+\mu+2T\times\ln\left[1+\exp\left(-\frac{\sqrt{\mu^2+\rho^2}}{T}\right)\right]\right\}. \tag{3.6}$$

This has the form of a Fermi quasiparticle density (for $\rho=0$ the expression (3.6) is simply the density of free fermions).

For the case $T=0$,[23,29] in which real time t replaces imaginary time τ, one can argue from Galilean invariance that the coefficient of $\partial_t\theta$ is rigorously related to the coefficient of $(\nabla\theta)^2$. It therefore does not appear in $\Omega^{(2)}_{\rm kin}$. We wish, however, to stress that these arguments cannot be used to eliminate the term $(\nabla\theta)^2$ from $\Omega^{(2)}_{\rm kin}$ when $T\neq 0$, so we must calculate it explicitly.

The O_1 term in Σ yields

$$\Omega_{\rm kin}^{(2)}(O_1)=\frac{T}{2}\int_0^\beta d\tau\int d\mathbf{r}-\frac{T}{(2\pi)^2}\times\sum_{n=-\infty}^{\infty}\int d\mathbf{k}\ {\rm Tr}[\mathscr{G}(i\omega_n,\mathbf{k})\tau_3\mathscr{G}(i\omega_n,\mathbf{k})\tau_3]\times\left(\frac{i\partial_\tau\theta}{2}+\frac{(\nabla\theta)^2}{8m}\right)^2. \tag{3.7}$$

Using (A11) to compute the sum over the Matsubara frequencies, we find that

$$\Omega_{\rm kin}^{(2)}(O_1)=-\frac{T}{2}\int_0^\beta d\tau\int d\mathbf{r}K(\mu,T,\rho)\left(i\partial_\tau\theta+\frac{(\nabla\theta)^2}{4m}\right)^2, \tag{3.8}$$

where

$$K[\mu,T,\rho(\mu,T)]=\frac{m}{8\pi}\left(1+\frac{\mu}{\sqrt{\mu^2+\rho^2}}\tanh\frac{\sqrt{\mu^2+\rho^2}}{2T}\right). \tag{3.9}$$

Obviously, the O_1 term does not affect the coefficient of $(\nabla\theta)^2$. Further, it is easy to make sure that the cross term involving O_1 and O_2 in $\Omega_{\rm kin}^{(2)}$ is absent. Finally, calculations of the O_2 contribution to $\Omega_{\rm kin}^{(2)}$ yield[4)]

$$\Omega_{\rm kin}^{(2)}(O_2)=T\int_0^\beta d\tau\int d\mathbf{r}\frac{T}{(2\pi)^2}\sum_{n=-\infty}^{\infty}\int d\mathbf{k}\,\mathbf{k}^2{\rm Tr}\times[\mathscr{G}(i\omega_n,\mathbf{k})\hat{I}\mathscr{G}(i\omega_n,\mathbf{k})\hat{I}]\frac{(\nabla\theta)^2}{16m^2}. \tag{3.10}$$

Thus, summing over the Matsubara frequencies [see Eq. (A12)], one obtains

$$\Omega_{\rm kin}^{(2)}(O_2)=-\int_0^\beta d\tau\int d\mathbf{r}\frac{1}{128\pi^2m^2}\times\int d\mathbf{k}\frac{\mathbf{k}^2}{\cosh^2\dfrac{\sqrt{\xi^2(\mathbf{k})+\rho^2}}{2T}}(\nabla\theta)^2. \tag{3.11}$$

As expected, this term vanishes when $T\to0$, but at finite T it is comparable with (3.5).

Combining (3.5), (3.11), and (3.8) we finally obtain

$$\Omega_{\rm kin}=\frac{T}{2}\int_0^\beta d\tau\int d\mathbf{r}[n_F(\mu,T,\rho)i\partial_\tau\theta+J(\mu,T,\rho)(\nabla\theta)^2+K(\mu,T,\rho)(\partial_\tau\theta)^2], \tag{3.12}$$

where

$$J[\mu,T,\rho(\mu,T)]=\frac{1}{4m}n_F(\mu,T,\rho)-\frac{T}{4\pi}\int_{-\mu/2T}^{\infty}dx\frac{x+\mu/2T}{\cosh^2\sqrt{x^2+\dfrac{\rho^2}{4T^2}}} \tag{3.13}$$

characterizes the phase stiffness and governs the spatial variation of the phase $\theta(\mathbf{r})$. One can see that our value of the phase stiffness $J(T=0)$ coincides with the nonrenormalized stiffness used in Ref. 27.

The quantity $J(\mu,T,\rho)$ vanishes at $\rho=0$, which means that above T_ρ the modulus-phase variables are meaningless; to study the model in this region one must use the old variables Φ and Φ^*. Near T_ρ one can obtain from (3.13) in the high-density limit (see below)

$$J(\mu\simeq\epsilon_F,T\to T_\rho,\rho\to0)=\frac{7\zeta(3)}{16\pi^3}\frac{\rho^2}{T_p^2}\varepsilon_F\simeq0.016\frac{\rho^2}{T_\rho^2}\epsilon_F. \tag{3.14}$$

Direct comparison of (3.12) with the Hamiltonian of the *XY*-model (3.1) makes it possible to write Eq. (3.2) for $T_{\rm BKT}$ directly:

$$\frac{\pi}{2}J[\mu,T_{\rm BKT},\rho(\mu,T_{\rm BKT})]=T_{\rm BKT}. \tag{3.15}$$

Although mathematically this reduces to a well-known problem, the analogy is incomplete. Indeed, in the standard *XY*-model (as well as the nonlinear σ-model) the vector (spin) subject to ordering is assumed to be a unit vector with no dependence on T.[5)] In our case this is definitely not the case, and a self-consistent calculation of $T_{\rm BKT}$ as a function of n_f requires additional equations for ρ and μ, which together with (3.15) form a complete set.

Using the definition (2.7), one can derive the effective potential $\Omega_{\rm pot}(v,\mu,T,\rho)$ (see Appendix A). Then the desired missing equations are the condition $\partial\Omega_{\rm pot}(\rho)/\partial\rho=0$ that the potential (A10) be minimized, and the equality $v^{-1}\partial\Omega_{\rm pot}/\partial\mu=-n_f$, which fixes n_f. These are, respectively

$$\frac{1}{V}=\int\frac{d\mathbf{k}}{(2\pi)^2}\frac{1}{2\sqrt{\xi^2(\mathbf{k})+\rho^2}}\tanh\frac{\sqrt{\xi^2(\mathbf{k})+\rho^2}}{2T}, \tag{3.16}$$

$$n_F(\mu,T,\rho)=n_f, \tag{3.17}$$

where $n_F(\mu,T,\rho)$ is defined by (3.16)

Equations (3.16) and (3.17) comprise a self-consistent system for determining the modulus ρ of the order parameter and the chemical potential μ in the mean-field approximation for fixed T and n_f.

While Eqs. (3.16) and (3.17) seem to yield a reasonable approximation at high densities n_f, since they include condensed boson pairs in a nonperturbative way via nonzero ρ, they must certainly be corrected in the strong coupling regime (low densities n_f) to take into account the contribution of noncondensed bosons (this appears to be important also for Eq. (3.15), which determines $T_{\rm BKT}$). The extent to which this alters the present results is not completely clear. Previously, the best way to incorporate noncondensed pairs seems to have been the self-consistent *T*-matrix approximation,[10,30–32] which allows one to account for the feedback of pairs on the self-energy of fermions. However, the *T*-matrix approach, at least in its standard form,[10,30–32] fails to describe the BKT phase transition, for which one must consider the equation for the vertex. On the other hand,

in our approach the BKT phase transition is realized by the condition (3.2), while an analog of the T-matrix approximation in terms of propagators of the ρ-particle and the neutral fermion χ has yet to be elaborated.

The energy of two-particle bound states in vacuum

$$\varepsilon_b = -2W\exp\left(-\frac{4\pi}{mV}\right) \tag{3.18}$$

(see Refs. 4, 7, and 33) is more convenient to use than the four-fermion constant V (here W is the conduction bandwidth). For example, one can easily take the limits $W\to\infty$ and $V\to 0$ in Eq. (3.16), which after this renormalization becomes

$$\ln\frac{|\varepsilon_b|}{\sqrt{\mu^2+\rho^2}-\mu} = 2\int_{-\mu/T}^{\infty} du \times \frac{1}{\sqrt{u^2+\left(\frac{\rho}{T}\right)^2}\left[\exp\sqrt{u^2+\left(\frac{\rho}{T}\right)^2}+1\right]}. \tag{3.19}$$

Thus, in practice, we solve Eqs. (3.15), (3.17), and (3.19) numerically to study $T_{\rm BKT}$ as function of n_f (or equivalently, of the Fermi energy $\epsilon_F=\pi n_f/m$, as it should be for 2D metals with the simplest quadratic dispersion law).

It is easy to show that at $T=0$, the system (3.17), (3.19) transforms into a previously studied system (see Ref. 4 and references therein). Its solution is $\rho=\sqrt{2|\varepsilon_b|\epsilon_F}$ and $\mu=-|\varepsilon_b|/2+\epsilon_F$. This will be useful in studying the concentration dependencies of $2\Delta/T_{\rm BKT}$ and $2\Delta/T_\rho$, where Δ is the zero-temperature gap in the quasiparticle excitation spectrum. It should be borne in mind that in the local pair regime ($\mu<0$), the gap Δ equals $\sqrt{\mu^2+\rho^2}$ rather than ρ (as in the case $\mu>0$).[4]

Setting $\rho=0$ in Eqs. (3.16) and (3.17), we obtain (in the same approximation) the equations for the critical temperature T_ρ and the corresponding value of μ:

$$\ln\frac{|\varepsilon_b|}{T_\rho}\frac{\gamma}{\pi} = -\int_0^{\mu/2T_\rho} du\,\frac{\tanh u}{u}\ (\gamma=1.781), \tag{3.20}$$

$$T_\rho \ln\left[1+\exp\left(\frac{\mu}{T_\rho}\right)\right] = \epsilon_F. \tag{3.21}$$

Note that these equations coincide with the system that determines the mean-field temperature $T_c^{(2D)MF}(=T_\rho)$ and $\mu(T_c^{(2D)MF})$,[7] evidently as a result of the mean-field approximation for the variable ρ used here. There is, however, an important difference between the temperatures T_c^{2D} and T_ρ. Specifically, if one takes fluctuations into account, T_c^{2D} goes to zero, while the value of T_ρ remains finite. The crucial point is that the perturbation theory in the variables ρ and θ does not contain any infrared singularities,[22,34] in contrast to the perturbation theory in Φ, Φ^*; thus the fluctuations do not reduce T_ρ to zero. This is why the temperature T_ρ has its own physical meaning: incoherent (local or Cooper) pairs begin to form (at least at high enough n_f [see Sec. 5)] just below T_ρ. At higher temperatures, only these pair fluctuations exist; their influence was studied in Ref. 35.

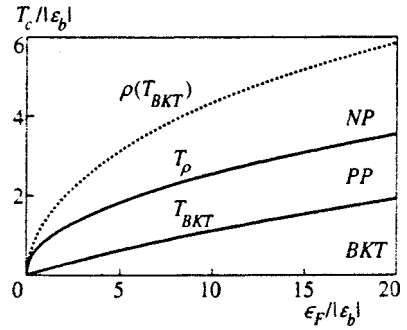

FIG. 1. $T_{\rm BKT}$ and T_ρ versus the noninteracting fermion density. Dots represent the function $\rho(\epsilon_F)$ at $T=T_{\rm BKT}$. The regions of normal phase (NP) pseudogap phase (PP), and BKT phase are indicated.

4. NUMERICAL RESULTS

A numerical investigation of the systems (3.15), (3.17), (3.19), and (3.20), (3.21) yields the following results, which are displayed graphically as the phase diagram of the system.

a) For low carrier densities, the pseudogap phase area (see Fig. 1) is comparable with the BKT area. For high carrier densities ($\epsilon_F \gtrsim 10^3|\varepsilon_b|$), one easily funds that the pseudogap region shrinks asymptotically as

$$\frac{T_\rho - T_{\rm BKT}}{T_\rho} \simeq \frac{4T_\rho}{\epsilon_F}. \tag{4.1}$$

This behavior qualitatively restores the BCS limit observed in overdoped samples.

b) For $\epsilon_F \leqslant (10-15)|\varepsilon_b|$, the function $T_{\rm BKT}(\epsilon_F)$ is linear, as also confirmed by the analytic solution of the system (3.15), (3.17), and (3.19), which yields $T_{\rm BKT}=\epsilon_F/8$. Remarkably, such a behavior of $T_c(\epsilon_F)$ is observed for all families of HTSC cuprates in their underdoped region,[3,27] though with a smaller coefficient of proportionality (0.01–0.1). This indicates the importance of including a contribution due to noncondensed pairs in Eq. (3.15), which defines $T_{\rm BKT}$.

It has been shown that for an optimal doping, the dimensionless ratio $\epsilon_F/|\varepsilon_b| \sim 3\times 10^2 - 10^3$.[36] Thus it is quite natural to suppose that in the underdoped region one has $\epsilon_F/|\varepsilon_b| \sim 10-10^2$, where we find linear behavior.

We note that in this limit, the temperature T_c of formation of a homogeneous order parameter for the quasi-2D model[3,6] can easily be written in the form

$$T_c \approx \frac{4T_{\rm BKT}}{\ln(\epsilon_F|\varepsilon_b|/4t_\parallel^2)}, \tag{4.2}$$

where $t_\parallel$ is the interplane hopping (coherent tunneling) constant. This shows that when $T_c < T_{\rm BKT}$, the weak three-dimensionalization can preserve (in any case, at low n_f) the regions of the pseudogap and BKT phases, which, for example, happens in the relativistic quasi-2D model.[34] At the same time, as the three-dimensionalization parameter $t_\parallel$ increases, when $T_c > T_{\rm BKT}$ the BKT phase can vanish, provided, however, that the anomalous phase region and both temperatures T_ρ and $T_c \simeq n_f/m$ are preserved.

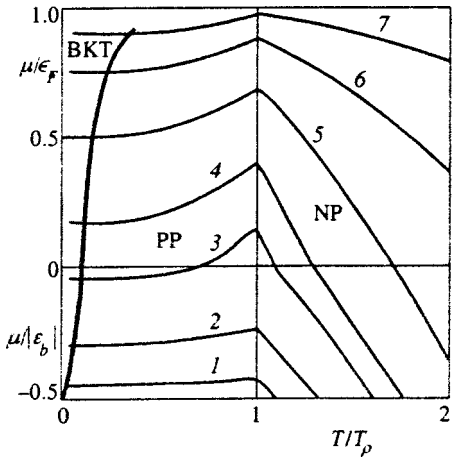

FIG. 2. $\mu(T)$ for various values of $\epsilon_F/|\varepsilon_b|$: *1*—0.05; *2*—0.2; *3*—0.45; *4*—0.6; *5*—1; *6*—2; *7*—5. (For $\mu>0$ and $\mu<0$ the chemical potential was scaled to ϵ_F and $|\varepsilon_b|$, respectively.) The thick lines delimit regions of BKT, pseudogap (PP), and normal (NP) phases.

c) Figure 2 shows the values of n_f for which μ differs substantially from ϵ_F, or in other words, the Landau Fermi-liquid theory becomes inapplicable to metals (also called bad metals) with low or intermediate carrier density. As expected, the kink μ at $T=T_\rho$, which has been observed experimentally[37] and interpreted for the 1-2-3 cuprates,[38] becomes less and less pronounced as ϵ_F increases. But in the present case it is interesting that in the hydrodynamic approximation employed here, it happens at the normal-pseudogap phase boundary or before superconductivity really appears. It would therefore be of great interest to perform experiments that might reveal the temperature dependence $\mu(T)$, especially for strongly anisotropic and relatively weakly doped cuprates.

d) It follows from curve *3* in Fig. 2 that the crossover (sign change in μ) from local to Cooper pairs is possible not only as ϵ_F increases, which is more or less obvious, but also (for some n_f) as T increases.

e) Finally the calculations showed (see Fig. 3) that the ratio $2\Delta/T_{\rm BKT}$ is greater than 4.7 in the region under study. The value $2\Delta/T_\rho(=2\Delta/T_c^{MF})$ is, however, somewhat lower and reaches the BCS theory limit of 3.52 only for $\epsilon_F \gg |\varepsilon_b|$. It is interesting that this concentration behavior is consistent with numerous measurements of this ratio in HTSC.[39,40] Note that the divergence of $2\Delta/T_{\rm BKT}$ and $2\Delta/T_\rho$ at $\epsilon_F\to 0$ is directly related to the definition of Δ at $\mu<0$.

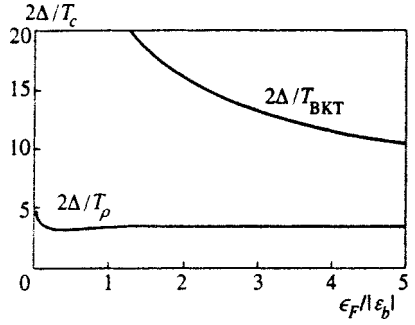

FIG. 3. $2\Delta/T_{\rm BKT}$ and $2\Delta/T_\rho$ versus the non-interacting fermion density.

5. PAIRING TEMPERATURE T_ρ VERSUS CARRIER DENSITY

There is no disagreement concerning the asymptotic behavior of $T_{\rm BKT}$ (or T_c) $\sim\epsilon_F$ in the region of low carrier densities. In contrast, the behavior of the temperature T_ρ, below which pairs are formed, cannot be considered to be generally accepted. For example, in Refs. 3 and, 27 based on qualitative arguments, this temperature is taken to be the temperature T_p of local uncorrelated pairing, which in contrast to T_ρ increases with decreasing n_f.[6)] Randeria (see Ref. 4 and references therein), to define the pairing temperature T_p, uses the system of equations for the mean-field transition temperature and the corresponding chemical potential, which is essentially identical to the system (3.20), (3.21). Thus his $T_p\to 0$ as $n_f\to 0$.

It is also well known[4,5,9] that in the low-density limit, it is vital to include quantum fluctuations, at least in the number equation,[12] in the calculation of the critical temperature at which a long-range order forms in 3D. In 2D these fluctuations in fact reduce the critical temperature to zero.[11] Certainly quantum fluctuations are also important in the calculation of T_ρ in the limit $n_f\to 0$ and, in particular, in the number equation. However, as already stressed in Sec. 3, these corrections are quite different from what we obtain using the variables Φ, Φ^*, since perturbation theory in the variables ρ and θ does not contain any infrared singularities,[22,34] and the fluctuations do not yield $T_\rho\equiv 0$. In fact, even including quantum fluctuations, T_ρ must exceed $T_{\rm BKT}$ $(\rho(T_{\rm BKT})\neq 0)$, so that the pseudogap phase is always present.

In our opinion, the temperature T_ρ has its own physical interpretation: this is the temperature of a smooth transition to the state in which the neutral order parameter $\rho\neq 0$, and below which one can observe pseudogap manifestations. There is also a very interesting and important question about the character of the transition. Certainly in the simplest Landau theory one appears to have a second-order phase transition, since ρ takes a nonzero value only below T_ρ.[14] However this kind of transition is only possible for neutral fermions. Fluctuations of the θ-phase will transform the pole in the Green function of the neutral fermions into a branch cut in the Green function for charged particles in the BKT phase. Indeed, the CMWH theorem concerning the absence of spontaneous breaking of a continuous symmetry means that symmetry-violating Green functions must vanish. However, it says nothing about the gap in the spectrum of excitations, as is sometimes incorrectly stated.

The correct explanation is that if the symmetry is unbroken, and the fermion excitation appears as a pole in the ψ two-point function, then the fermion must be gapless. If the fermion does not have the same quantum numbers as ψ (like our fermion χ) and so does not appear in the ψ two-point function as a one-particle state, then the symmetry does not tell whether the fermion (χ) will be gapless or not.

This very general argument[22] suggests the following plausible scenario. At low temperatures ($T<T_{\rm BKT}$), χ, ρ, and

θ should be treated as physical quasiparticles (χ, ρ having a gap and θ being a gapless excitation), while a straightforward computation of the ψ two-point function[22] reveals its branch-cut structure.

On the other hand, at temperatures above $T_{\rm BKT}$, we should consider ψ and Φ true quasiparticles, since $T_{\rm BKT}$ is a phase transition point and the spectrum of physical excitations changes precisely at this point. The ψ two-point function at $T>T_{\rm BKT}$ should be studied separately due to the presense of vortices which change the form of the correlator $\langle\exp[i\theta(x)]\exp[i\theta(0)]\rangle$ above $T_{\rm BKT}$. In this temperature region the ψ two-point function loses its branch-cut structure; instead, it acquires the form suggested in Refs. 30 and 31 with a pseudogap originating from the superconducting gap below $T_{\rm BKT}$, which preserves ''BCS-like'' structure as well as the diagonal component of the single-particle Green function. In this picture the Fermi-liquid description breaks down, evidently below T_ρ, due to the formation of nonzero ρ.

We note, however, that the decisive confirmation of this picture demands further detailed study probably based on a different approach, for example the self-consistent T-matrix (see Ref. 30 and references therein), which enables one to directly obtain the full fermion Green function.

To define the temperature T_p properly, one should study the spectrum of bound states either by solving the Bethe–Salpeter equation[7] or by analyzing the corresponding Green functions as we do here. It turns out that there is no difference between T_p and T_ρ in the Cooper pair regime ($\mu>0$), while in the local pair region ($\mu<0$) these temperatures exhibit different behavior.

Indeed, let us study the spectrum of bound states in both the normal ($\rho=0$) and pseudogap ($\rho\neq0$) phases. We are especially interested in determining the conditions under which real bound states (with zero total momentum $\mathbf{K}=0$) become unstable. For this purpose one can look at the propagator of the ρ-particle in the pseudogap phase:

$$\Gamma^{-1}(\tau,\mathbf{r})=\frac{1}{2}\left.\frac{\beta\delta^2\Omega(v,\mu,T,\rho(\tau,\mathbf{r}),\partial\theta(\tau,\mathbf{r}))}{\delta\rho(\tau,\mathbf{r})\,\delta\rho(0,0)}\right|_{\rho=\rho_{\min}=\text{const}}, \tag{5.1}$$

where $\rho_{\min}$ is defined by the minimum condition (3.16) [or (3.19)] of the potential part (A10) of the effective action (2.3). In the momentum representation, the spectrum of bound states is usually determined by the condition

$$\Gamma_R^{-1}(\omega,\mathbf{K})=0, \tag{5.2}$$

where $\Gamma_R(\omega,\mathbf{K})$ is the retarded Green function obtained directly from the temperature Green function $\Gamma(i\Omega_n,\mathbf{K})$ using the analytic continuation $i\Omega_n\to\omega+i0$. Recall that such an analytic continuation must be performed after evaluating the sum over the Matsubara frequencies. In case of vanishing total momentum $\mathbf{K}=0$, one arrives at the energy spectrum equation

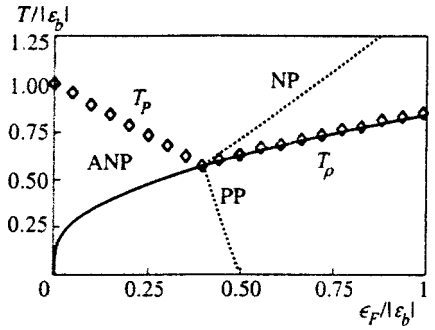

FIG. 4. Phase diagram of the 2D-metal at low concentrations. The dotted line corresponds to $\mu=0$, and the temperature T_p separates abnormal normal phase ANP from normal phase. The critical temperature $T_{\rm BKT}$ is not shown.

$$\Gamma_R^{-1}(\omega,0)=\frac{1}{V}+2\int\frac{d\mathbf{k}}{(2\pi)^2}\frac{\xi^2(\mathbf{k})}{\sqrt{\xi^2(\mathbf{k})+\rho^2}}\times\frac{\tanh\sqrt{\xi^2(\mathbf{k})+\rho^2}/2T}{\omega^2-4[\xi^2(\mathbf{k})+\rho^2]}=0. \tag{5.3}$$

From the explicit expression (5.3) for $\Gamma_R(\omega,0)$, this function obviously has a branch cut at frequencies

$$|\omega|\geqslant 2\min\sqrt{\xi^2(\mathbf{k})+\rho^2}=\begin{cases}2\rho, & \mu\geqslant 0,\\ 2\sqrt{\mu^2+\rho^2}, & \mu<0.\end{cases} \tag{5.4}$$

Thus, bound states can exist below this cut.

Real bound states decay into two-fermion states when the energy of the former reaches the branch point $2\min\sqrt{\xi^2(\mathbf{k})+\rho^2}$. Since Γ_R^{-1} is a monotonously decreasing function of ω^2, it has the unique solution $|\varepsilon_b(T)|=2\rho(T)$, at which Eq. (5.3) coincides exactly with the mean-field equation (3.16) for $\rho(T)$. It also becomes clear that for $\mu<0$ we have real bound states with energy $\varepsilon_b(T)$ below the two-particle scattering continuum at $\omega=2\sqrt{\mu^2+\rho^2}$, while at $\mu\geqslant0$ there are no stable bound states. The line $\mu(T,\epsilon_F)=0$ in the $T-\epsilon_F$ plane at $\rho\neq0$ separates the negative μ region where local pairs exist from that in which only Cooper pairs exist (positive μ). This line (see Fig. 4) begins at the point $T=(e^\gamma/\pi)|\varepsilon_b|\approx0.6|\varepsilon_b|, \epsilon_F\approx0.39|\varepsilon_b|$ and ends at $T=0$, $\epsilon_F=|\varepsilon_b|/2$. (The latter follows directly from the solution at $T=0$, $\mu=-|\varepsilon_b|/2+\epsilon_F$[4,7].)

To find a similar line in the normal phase with $\rho=0$, we consider the corresponding equation for the bound states. The propagator of these states (in imaginary time formalism) is defined to be

$$\Gamma^{-1}(\tau,\mathbf{r})=\left.\frac{\beta\delta^2\Omega(v,\mu,T,\Phi(\tau,\mathbf{r}),\Phi^*(\tau,\mathbf{r}))}{\delta\Phi^*(\tau,\mathbf{r})\,\delta\Phi(0,0)}\right|_{\Phi=\Phi^*=0}. \tag{5.5}$$

[In the normal phase, where $\rho=0$, we must again use the initial auxiliary fields Φ and Φ^* (see Secs. 2 and 3)]. Then in the momentum representation (after summing over the Matsubara frequencies) we have

$$\Gamma^{-1}(i\Omega_n,\mathbf{K})=\frac{1}{V}-\frac{1}{2}\int\frac{d\mathbf{k}}{(2\pi)^2}\times\frac{\tanh\xi_+(\mathbf{k},\mathbf{K})/2T+\tanh\xi_-(\mathbf{k},\mathbf{K})/2T}{\xi_+(\mathbf{k},\mathbf{K})+\xi_-(\mathbf{k},\mathbf{K})-i\Omega_n}\times\xi_\pm(\mathbf{k},\mathbf{K})\equiv\frac{1}{2m}\left(\mathbf{k}\pm\frac{\mathbf{K}}{2}\right)^2-\mu, \quad (5.6)$$

where $\mathbf{k}$ is the relative momentum of the pair. The spectrum of bound states is given again by Eq. (5.2). Using the energy ε_b [see Eq. (3.18)] of the bound state at $T=0$, for $\mathbf{K}=0$ we obtain the following equation for the energies of these states in the normal phase:

$$\int_0^\infty dx\left[\frac{1}{x^+|\varepsilon_b|/2}-\frac{\tanh(x-\mu)/2T}{x-\mu-\omega/2}\right]=0. \quad (5.7)$$

Such states can exist provided $-2\mu-|\varepsilon_b|<\omega<-2\mu$. The left-hand side of Eq. (5.7) is positive at $\omega=-2\mu-|\varepsilon_b|$ and tends to $+\infty(\mu>0)$ or $-\infty(\mu<0)$ when $\omega\to-2\mu$. This equation always has a solution at $\mu<0$, so bound states with zero total momentum exist for negative μ.

For $\mu>0$, analytic analysis becomes more complicated, and requires numerical study. One can easily find from (5.7) that at $T=0$, stable bound states exist up to $\mu<|\varepsilon_b|/8$. In fact, numerical study for $T\geqslant T_\rho$ shows that the trajectory $\mu(T,\epsilon_F)=0$ [or $T=\epsilon_F/\ln 2$, see (3.21)] approximately divides the normal phase into two qualitatively different regions: with ($\mu<0$) and without ($\mu>0$) stable (long-lived) pairs. This also holds for other phases, which enables one to draw the whole line $\mu(T,\epsilon_F)=0$ (Fig. 4).

Knowing the two-particle binding energy, it is natural to define pairing temperature T_p as $T_p\approx|\varepsilon_b[T_p,\mu(T_p,\epsilon_F)]$. This equation can be easily analyzed in the region $\epsilon_F\ll|\varepsilon_b|$, for which we directly obtain $T_p\approx|\varepsilon_b|$, which clearly coincides with the standard estimate.[3,41] This means in turn that the curve $T_p(\epsilon_F)$ starting at $T_p(0)\approx|\varepsilon_b|$ will be reduced, up to the point $T_p(0.39\epsilon_F)\approx0.6|\varepsilon_b|$, which lies on the line $T_\rho(\epsilon_F)$ (see Fig. 4). It is important that this line is not the phase transition curve; it merely divides the fermion system diagram into temperature regions with a prevailing mean number of local pairs ($T\lesssim T_p$) or unbound carriers ($T\gtrsim T_p$). This is the region of the abnormal normal phase where one has preformed boson pairs. It is widely accepted, however, that this case is only of theoretical interest, since there is no Fermi surface ($\mu<0$) in the phase. The phase area or the difference $T_p(\epsilon_F)-T_\rho(\epsilon_F)$ is an increasing function as $\epsilon_F\to0$, which corresponds to the behavior usually assumed.[3,27]

When $\mu>0$ there are no stable bound states ($\varepsilon_b(T)=2\rho(T)=0$) for the normal phase, where they are short-lived. Formally, using $\rho(T)=0$ in Eq. (5.3), we immediately obtain (3.20) or, in other words, here $T_p=T_\rho$. Such a conclusion is in accordance with the generally accepted definition of T_p in the BCS case.[41]

Thus the phase diagram of a 2D metal above T_c acquires the form shown in Fig. 4. It is interesting that if the line $T_p(\epsilon_F)$ cannot be defined exactly, the temperature $T_\rho(\epsilon_F)$ is the line below which pairs reveal some signs of collective behavior. Moreover, at $T<T_\rho$ one can speak of a real pseudogap in the one-particle spectrum, while in the region $T_\rho<T<T_p$ only strongly developed pair fluctuations (some number of pairs) exist, though they probably suffice to reduce the spectral quasi-particle weight, and to produce other observed manifestations that mask pseudogap (spin gap; see Ref. 35) formation.

6. PARAMAGNETIC SUSCEPTIBILITY OF THE SYSTEM

It would be very interesting to study how a nonzero value of the neutral order parameter affects the observable properties of the 2D system. Does this really resemble the gap opening in the traditional superconductors, except that it happens in the normal phase? Or, in other words, does the pseudogap open?

We shall demonstrate this phenomenon, taking the paramagnetic susceptibility of the system as the simplest case in point. To study the system in the magnetic field $\mathbf{H}$ one must add the paramagnetic term

$$\mathscr{H}_{PM}=-\mu_B H[\psi_\uparrow^\dagger(\mathbf{r})\psi_\uparrow(\mathbf{r})-\psi_\downarrow^\dagger(\mathbf{r})\psi_\downarrow(\mathbf{r})] \quad (6.1)$$

to the Hamiltonian (2.1) where $\mu_B=e\hbar/2mc$ is the Bohr magneton. Note that, using the isotropy in the problem, we chose the direction of field $\mathbf{H}$ to be perpendicular to the plane containing the vectors $\mathbf{r}$.

Adding the corresponding term to Eq. (2.8) for the neutral fermion Green function, it is easy to show that in the momentum representation [compare with (3.4)]

$$\mathscr{G}(i\omega_n,\mathbf{k},H)=\frac{(i\omega_n+\mu_B H)\hat{I}+\tau_3\xi(\mathbf{k})-\tau_1\rho}{(i\omega_n+\mu_B H)^2-\xi^2(\mathbf{k})-\rho^2}. \quad (6.2)$$

The static paramagnetic susceptibility can be expressed in terms of the magnetization,

$$\chi(\mu,T,\rho)=\left.\frac{\partial M(\mu,T,\rho,H)}{\partial H}\right|_{H=0}, \quad (6.3)$$

which in the mean-field approximation can be derived from the effective potential:

$$M(\mu,T,\rho,H)=-\frac{1}{v}\frac{\partial\Omega_{\rm pot}(v,\mu,T,\rho,H)}{\partial H}. \quad (6.4)$$

Thus from (6.4) one obtains

$$M(\mu,T,\rho,H)=\mu_B T\sum_{n=-\infty}^{\infty}\int\frac{d\mathbf{k}}{(2\pi)^2}\mathrm{Tr}[\mathscr{G}(i\omega_n,\mathbf{k},H)\hat{I}]. \quad (6.5)$$

Then using the definition (6.3) one arrives at

$$\chi(\mu,T,\rho)=\mu_B^2\int\frac{d\mathbf{k}}{(2\pi)^2}2T\sum_{n=-\infty}^{\infty}\frac{\xi^2(\mathbf{k})+\rho^2-\omega_n^2}{[\omega_n^2+\xi^2(\mathbf{k})+\rho^2]^2}. \quad (6.6)$$

The sum in (6.6) can easily be calculated with the help of Eq. (A11); thus, we obtain the final result

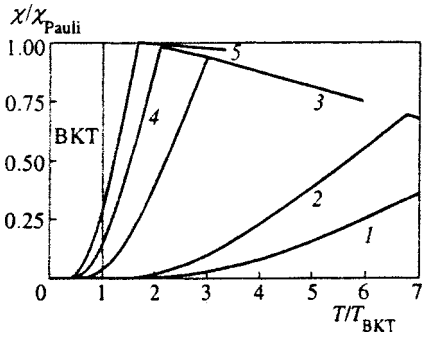

FIG. 5. $\chi(T)$ for various values of $\epsilon_F/|\varepsilon_b|$: *1*—0.6; *2*—1; *3*—5; *4*—10; *5*—30.

$$\chi(\mu,T,\rho)=\chi_{\text{Pauli}}\frac{1}{2}\int_{-\mu/2T}^{\infty}\frac{dx}{\cosh^2\sqrt{x^2+\dfrac{\rho^2}{4T^2}}}, \tag{6.7}$$

where $\chi_{\text{Pauli}}\equiv\mu_B^2 m/\pi$ is the Pauli paramagnetic susceptibility for the 2D system.

To study χ as a function of T and n_f (or ϵ_F), Eq. (6.7) should be used together with Eqs. (3.17) and (3.19).

For the case of the normal phase ($\rho=0$) one can investigate the system analytically. Thus (6.7) takes the form

$$\chi(\mu,T,\rho=0)=\chi_{\text{Pauli}}\frac{1}{1+\exp(-\mu/T)}, \tag{6.8}$$

where μ is determined by (3.21). This system has the solution

$$\chi(\epsilon_F,T,\rho=0)=\chi_{\text{Pauli}}\,[1-\exp(-\epsilon_F/T)], \tag{6.9}$$

which is identical to a solution known from the literature.[42]

The results of a numerical study of the system (6.7), (3.19), and (3.17) are presented in Fig. 5. One can see that the kink in χ occurs at $T=T_\rho$ as in the dependence of μ on T. Below T_ρ the value of $\chi(T)$ decreases, although the system is still normal. This can be interpreted as a spin-gap (or pseudogap) opening. The size of the pseudogap region depends strongly on the doping ($\epsilon_F/|\varepsilon_b|$), as observed for real HTSC.[2,18,19] For small values of $\epsilon_F/|\varepsilon_b|$ this region is very large ($T_\rho>6T_{\text{BKT}}$), while for large $\epsilon_F/|\varepsilon_b|\sim 5-30$ it is slightly larger than the region corresponding to the BKT phase.

7. CONCLUSION

To summarize, we have discussed the crossover in the superconducting transition between BCS- and Bose-like behavior for the simplest 2D model, with *s*-wave nonretarded attractive interaction.

While there is still no generally accepted microscopic theory of HTSC compounds and their basic features (including the pairing mechanism), it seems that this approach, although in a sense phenomenological, is of great interest since it is able to cover the whole range of carrier concentrations (and thus the whole range of coupling constants) and temperatures. As we tried to demonstrate, it enables one to propose both a reasonable interpretation for the observed phenomena caused by doping and to describe new phenomena; for example, pseudogap phase formation as a new thermodynamically equilibrium normal state of low-dimensional conducting electronic systems.

Evidently there are a number of important open questions. They may be divided into two classes: the first concerns the problem of a better and more complete treatment of the models themselves. The second class relates to the extent to which this model is applicable to HTSC compounds, and what the necessary ingredients are for a more realistic description.

Regarding the microscopic Hamiltonian as a given model, our treatment is obviously still incomplete. In particular, there exists an unconfirmed numerical result[43] based on a fully self-consistent determination of a phase transition to a superconducting state in a conserving approximation, which states that the superconducting transition is neither the simple mean-field transition nor the BKT transition. [See, however, the discussion preceding Eq. (3.18)] Besides, it would be very interesting to obtain the spectrum of both the anomalous normal and pseudogap phases. It is important also to take into consideration the effects of noncondensed bosons, which might help to obtain a smaller slope in the dependence of T_{BKT} on ϵ_F.

As for the extent to which the models considered are really applicable to HTSC, most of the complexity of these systems is obviously neglected here. For example, we did not take into account the indirect nature of attraction between the fermions, *d*-wave pairing, inter-layer tunneling, etc. Nevertheless, one may hope that the present simple model can explain the essential features of pseudogap formation.

We thank Drs. E. V. Gorbar, I. A. Shovkovy, O. Tchernyshyov, and V. M. Turkowski for fruitful discussions, which helped to clarify some deep questions about low-dimensional phase transitions. We especially thank Prof. R. M. Quick for many thoughtful comments on an earlier version of this manuscript. One of us (S. G. S.) is grateful to the members of the Department of Physics of the University of Pretoria, especially Prof. R. M. Quick and Dr. N. J. Davidson, for very useful points and hospitality. S. G. S. also acknowledges the financial support of the Foundation for Research Development, Pretoria.

APPENDIX A

Calculation of the effective potential

Here we sketch the derivation of the effective potential. To obtain it one must write Eq. (2.7) in the momentum representation:

$$\begin{aligned}\Omega_{\text{pot}}(v,\mu,T,\rho)&=v\left\{\frac{\rho^2}{V}-T\sum_{n=-\infty}^{+\infty}\int\frac{d\mathbf{k}}{(2\pi)^2}\text{Tr}[\ln\mathscr{G}^{-1}(i\omega_n,\mathbf{k})e^{i\delta\omega_n\tau_3}]\right.\\&\left.+T\sum_{n=-\infty}^{+\infty}\int\frac{d\mathbf{k}}{(2\pi)^2}\text{Tr}[\ln G_0^{-1}(i\omega_n,\mathbf{k})e^{i\delta\omega_n\tau_3}]\right\},\\&\delta\to+0,\end{aligned} \tag{A1}$$

where

$$\mathscr{G}^{-1}(i\omega_n,\mathbf{k})=i\omega_n\hat{I}-\tau_3\xi(\mathbf{k})+\tau_1\rho,$$
$$G_0^{-1}(i\omega_n,\mathbf{k})=\mathscr{G}^{-1}(i\omega_n,\mathbf{k})|_{\rho=\mu=0} \tag{A2}$$

are the inverse Green functions. The exponential factor $e^{i\delta\omega_n\tau_3}$ is added to (A1) to provide the correct regularization which is necessary to perform the calculation with the Green functions.[44] For instance, one obtains

$$\lim_{\delta\to+0}\sum_{n=-\infty}^{+\infty}\mathrm{Tr}[\ln\mathscr{G}^{-1}(i\omega_n,\mathbf{k})e^{i\delta\omega_n\tau_3}]$$
$$=\lim_{\delta\to+0}\left\{\sum_{n=-\infty}^{+\infty}\mathrm{Tr}[\ln\mathscr{G}^{-1}(i\omega_n,\mathbf{k})]\cos\delta\omega_n\right.$$
$$+i\sum_{\omega_n>0}\sin\delta\omega_n\mathrm{Tr}[(\ln\mathscr{G}^{-1}(i\omega_n,\mathbf{k})$$
$$\left.-\ln\mathscr{G}^{-1}(-i\omega_n,\mathbf{k}))\tau_3]\right\}$$
$$=\sum_{n=-\infty}^{+\infty}\mathrm{Tr}[\ln\mathscr{G}^{-1}(i\omega_n,\mathbf{k})]-\frac{\xi(\mathbf{k})}{T}, \tag{A3}$$

where

$$\ln\frac{\mathscr{G}^{-1}(i\omega_n,\mathbf{k})}{i\omega_n}\simeq\frac{-\tau_3\xi(\mathbf{k})+\tau_1\rho}{i\omega_n},\quad \omega_n\to\infty$$

and

$$\sum_{\omega_n>0}\frac{\sin\delta\omega_n}{\omega_n}\simeq\frac{1}{2\pi T}\int_0^\infty dx\frac{\sin\delta x}{x}=\frac{1}{4T}\mathrm{sign}\,\delta.$$

To calculate the sum in (A3), one must first use the identity $\mathrm{Tr}\ln\hat{A}=\ln\det\hat{A}$, so that (A1) takes the form

$$\Omega_{\mathrm{pot}}(v,\mu,T,\rho)=v\left\{\frac{\rho^2}{V}-T\sum_{n=-\infty}^{+\infty}\right.$$
$$\times\int\frac{d\mathbf{k}}{(2\pi)^2}\ln\frac{\det\mathscr{G}^{-1}(i\omega_n,\mathbf{k})}{\det G_0^{-1}(i\omega_n,\mathbf{k})}$$
$$\left.-\int\frac{d\mathbf{k}}{(2\pi)^2}[-\xi(\mathbf{k})+\varepsilon(\mathbf{k})]\right\}. \tag{A4}$$

Calculating the determinants of the Green functions (A2), one obtains

$$\Omega_{\mathrm{pot}}(v,\mu,T,\rho)=v\left\{\frac{\rho^2}{V}-T\right.$$
$$\times\sum_{n=-\infty}^{+\infty}\int\frac{d\mathbf{k}}{(2\pi)^2}\ln\frac{\omega_n^2+\xi^2(\mathbf{k})+\rho^2}{\omega_n^2+\varepsilon^2(\mathbf{k})}$$
$$\left.-\int\frac{d\mathbf{k}}{(2\pi)^2}[-\xi(\mathbf{k})+\varepsilon(\mathbf{k})]\right\}, \tag{A5}$$

where the role of $G_0(i\omega_n,\mathbf{k})$ in the regularization of Ω_{pot} is now evident. The summation in (A5) can be done if one uses the representation

$$\ln\frac{\omega_n^2+a^2}{\omega_n^2+b^2}=\int_0^\infty dx\left(\frac{1}{\omega_n^2+a^2+x}-\frac{1}{\omega_n^2+b^2+x}\right), \tag{A6}$$

and then

$$\sum_{k=0}^{\infty}\frac{1}{(2k+1)^2+c^2}=\frac{\pi}{4c}\tanh\frac{\pi c}{2}. \tag{A7}$$

We find

$$\ln\frac{\omega_n^2+a^2}{\omega_n^2+b^2}=\int_0^\infty dx\left(\frac{1}{2\sqrt{b^2+x}}\tanh\frac{\sqrt{b^2+x}}{2T}\right.$$
$$\left.-\frac{1}{2\sqrt{a^2+x}}\tanh\frac{\sqrt{a^2+x}}{2T}\right). \tag{A8}$$

Integrating (A8) over x, one thus obtains

$$T\sum_{n=-\infty}^{+\infty}\int\frac{d\mathbf{k}}{(2\pi)^2}\ln\frac{\omega_n^2+\xi^2(\mathbf{k})+\rho^2}{\omega_n^2+\varepsilon^2(\mathbf{k})}$$
$$=2T\int\frac{d\mathbf{k}}{(2\pi)^2}\ln\frac{\cosh[\sqrt{\xi^2(\mathbf{k})+\rho^2}/2T]}{\cosh[\varepsilon(\mathbf{k})/2T]}. \tag{A9}$$

Finally, substituting (A9) into (A5),

$$\Omega_{\mathrm{pot}}(v,\mu,T,\rho)=v\left\{\frac{\rho^2}{V}-\int\frac{d\mathbf{k}}{(2\pi)^2}\left[2T\ln\right.\right.$$
$$\times\frac{\cosh[\sqrt{\xi^2(\mathbf{k})+\rho^2}/2T]}{\cosh[\varepsilon(\mathbf{k})/2T]}$$
$$\left.\left.-[\xi(\mathbf{k})-\varepsilon(\mathbf{k})]\right]\right\}. \tag{A10}$$

It is easy to show that at $T=0$, the expression (A10) reduces to that obtained in Ref. 7.

Finally, we give formulas for the summation over the Matsubara frequencies used in Secs. 3 and 6:

$$T\sum_{n=-\infty}^{\infty}\mathrm{Tr}[\mathscr{G}(i\omega_n,\mathbf{k})\tau_3\mathscr{G}(i\omega_n,\mathbf{k})\tau_3]$$
$$=2T\sum_{n=-\infty}^{\infty}\frac{\xi^2(\mathbf{k})-\rho^2-\omega_n^2}{[\omega_n^2+\xi^2(\mathbf{k})+\rho^2]^2}$$
$$=-\frac{\rho^2}{[\xi^2(\mathbf{k})+\rho^2]^{3/2}}\tanh\frac{\sqrt{\xi^2(\mathbf{k})+\rho^2}}{2T}$$
$$-\frac{\xi^2(\mathbf{k})}{2T[\xi^2(\mathbf{k})+\rho^2]}\frac{1}{\cosh^2\dfrac{\sqrt{\xi^2(\mathbf{k})+\rho^2}}{2T}}, \tag{A11}$$

$$T\sum_{n=-\infty}^{\infty}\mathrm{Tr}[\mathscr{G}(i\omega_n,\mathbf{k})\hat{I}\mathscr{G}(i\omega_n,\mathbf{k})\hat{I}]$$
$$=2T\sum_{n=-\infty}^{\infty}\frac{\xi^2(\mathbf{k})+\rho^2-\omega_n^2}{[\omega_n^2+\xi^2(\mathbf{k})+\rho^2]^2}$$
$$=-\frac{1}{2T}\left(\cosh^2\frac{\sqrt{\xi^2(\mathbf{k})+\rho^2}}{2T}\right)^{-1}, \qquad \text{(A12)}$$

where the Green function $\mathscr{G}(i\omega_n,\mathbf{k})$ is given by (3.4). Both formulas can easily be calculated using Eq. (A7) and its derivative with respect to c.

*)E-mail: vgusynin@bitp.kiev.ua
†)E-mail: vloktev@bitp.kiev.ua

1)To calculate the observed single-particle spectrum, of course, carrier losses due to scattering of carriers by fluctuations of the phase of the order parameter (and in real systems by dopants) must be taken into account; see Ref. 17.

2)We note that the specific heat experiments[2] demonstrated the loss of entropy that occurs at temperatures much higher than T_c. This can be considered indicative of a degenerate normal state, consistent with the existence of a nonzero order parameter $\langle\rho\rangle$.

3)An exponentially small correction is omitted here.

4)Derivatives higher than $(\nabla\theta)^2$ were not computed here.

5)There is no doubt that in certain situations (for example, very high T) it also can become a thermodynamic variable, i.e., one dependent on T, as happens in problems of phase transitions between ordered (magnetic) and disordered (paramagnetic) phases when the spin itself vanishes. Specifically, for quasi-2D spin systems it is obvious that as one proceeds from high-T regions, a spin modulus first forms in 2D clusters of finite size and only then does global 3D ordering occur. We note, however, that this dependence was neglected in Ref. 27, where the nonrenormalized phase stiffness $J(T=0)$ was used to write Eq. (3.15).

6)In fact, in Refs. 3 and 27 (see also Ref. 5) this temperature was plotted as an increasing function of coupling constant V, which for 3D systems corresponds, to some extent, to the carrier density decreasing. In 2D systems, however, where, as is well known, two-particle bound states are formed without any threshold, similar conclusions about the behavior of $T_p(n_f)$ are questionable, and must be checked independently.

[1] V. M. Loktev, Fiz. Nizk. Temp. **22**, 3 (1996) [Low Temp. Phys. **22**, 1 (1996)].
[2] M. Randeria, Varenna Lectures, 1997, e-print archiv, cond-mat/97 10 223.
[3] Y. J. Uemura, e-print archiv, cond-mat/97 06 151, submitted in Physica C.
[4] M. Randeria, in *Bose Einstein Condensation*, ed. by A. Griffin, D. W. Snoke, and S. Stringari, Cambridge U. P., New York (1995) p. 355.
[5] R. Haussmann, Phys. Rev. B **49**, 12975 (1994).
[6] E. V. Gorbar, V. M. Loktev, and S. G. Sharapov, Physica C **257**, 355 (1996).
[7] E. V. Gorbar, V. P. Gusynin, and V. M. Loktev, Fiz. Nizk. Temp. **19**, 1171 (1993) [Low Temp. Phys. **19**, 832 (1993)]; Supercond., Phys. Chem. Technol. **6**, 375 (1993); Preprint ITP-92-54E, Kiev (1992).
[8] N. D. Mermin and H. Wagner, Phys. Rev. Lett. **17**, 1113 (1966); P. C. Hohenberg, Phys. Rev. **158**, 383 (1967); S. Coleman, Commun. Math. Phys. **31**, 259 (1973).
[9] S. Schmitt-Rink, C. M. Varma, and A. E. Ruckenstein, Phys. Rev. Lett. **63**, 445 (1989).
[10] J. Serene, Phys. Rev. B **40**, 10873 (1989).
[11] A. Tokumitu, K. Miyake, and K. Yamada, Phys. Rev. B **47**, 11988 (1993).
[12] P. Nozières and S. Schmitt-Rink, J. Low Temp. Phys. **59**, 195 (1985).
[13] S. V. Traven, Phys. Rev. Lett. **73**, 3451 (1994).
[14] R. MacKenzie, P. K. Panigrahi, and S. Sakhi, Int. J. Mod. Phys. A **9**, 3603 (1994); Phys. Rev. B **48**, 3892 (1993).
[15] M. Drechsler and W. Zwerger, Ann. Phys. (Germany) **1**, 15 (1992).
[16] S. Stintzing and W. Zwerger, Phys. Rev. B **56**, 9004 (1997).
[17] V. M. Loktev and Yu. G. Pogorelov, Physica C **272**, 151 (1996).
[18] D. Pines, J. of Physics **20**, 535 (1996).
[19] H. Ding, T. Yokoya, I. C. Campuzano *et al.*, Nature (London) **382**, 51 (1996).
[20] V. P. Gusynin, V. M. Loktev, and S. G. Sharapov, JETP Lett. **65**, 182 (1997); Fiz. Nizk. Temp. **23**, 1247 (1997) [Low Temp. Phys. **23**, 936 (1997)].
[21] H. Kleinert, Fortschr. Phys. **26**, 565 (1978).
[22] E. Witten, Nucl. Phys. B **145**, 110 (1978).
[23] I. J. R. Aitchison, P. Ao, D. J. Thouless, and X.-M. Zhu, Phys. Rev. B **51**, 6531 (1995).
[24] M. Capezzali, D. Ariosa, and H. Beck, Physica B **230-232**, 962 (1997).
[25] Yu. A. Izyumov and Yu. N. Skryabin, *Statistical Mechanics of Magnetically Ordered Systems*, [in Russian], Nauka, Moscow (1987), Ch. 15.
[26] P. Minnhagen, Rev. Mod. Phys. **59**, 1001 (1987).
[27] V. Emery and S. A. Kivelson, Nature (London) **374**, 434 (1995); Phys. Rev. Lett. **74**, 3253 (1995); e-print archiv, cond-mat/97 10 059.
[28] S. Doniach and M. Inui, Phys. Rev. B **41**, 6668 (1990).
[29] A. M. J. Schakel, Mod. Phys. Lett. B **4**, 927 (1990).
[30] O. Tchernyshyov, Phys. Rev. B **56**, 3372 (1997).
[31] B. Janko, J. Mali, and K. Levin, Phys. Rev. B **56**, 11407 (1997).
[32] J. R. Engelbrecht, A. Nazarenko, M. Randeria, and E. Dagotto, e-print archiv, cond-mat/97 05 166.
[33] K. Miyake, Prog. Theor. Phys. **69**, 1794 (1983).
[34] H. Yamamoto and I. Ichinose, Nucl. Phys. B **370**, 695 (1992).
[35] V. M. Loktev and S. G. Sharapov, Fiz. Nizk. Temp. **23**, 180 (1997) [Low Temp. Phys. **23**, 132 (1997)].
[36] M. Casas, J. M. Getino, M. de Llano *et al.*, Phys. Rev. B **50**, 15945 (1994).
[37] D. van der Marel, Physica C **165**, 35 (1990).
[38] A. V. Dotsenko and O. P. Sushkov, e-print archiv, cond-mat/96 01 031.
[39] T. P. Devereaux, Phys. Rev. Lett. **74**, 4313 (1995).
[40] C. Kendziora *et al.*, Phys. Rev. Lett. **77**, 727 (1996).
[41] R. Micnas and T. Kostyrko, in *Proceedings of the 1st Polish-US Conference*, Wroclaw, Poland, 11–15 September (1995).
[42] I. A. Kvasnikov, *Thermodynamics and Statistical Physics* [in Russian], Moscow University Press, Moscow (1991).
[43] J. J. Deisz, D. W. Hess, and J. W. Serene, e-print archiv, cond-mat/97 06 012 .
[44] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics*, Dover, New York (1975).

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

Strong coupling in the Kondo problem in the low-temperature region

Yu. N. Ovchinnikov and A. M. Dyugaev

L. D. Landau Institute for Theoretical Physics, 117940 Moscow, Russia
(Submitted 13 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1263–1284 (April 1999)

The magnetic field dependence of the average spin of a localized electron coupled to conduction electrons with an antiferromangetic exchange interaction is found for the ground state. In the magnetic field range $\mu H \sim 0.5 T_c$ (T_c is the Kondo temperature) there is an inflection point, and in the strong magnetic field range $\mu H \gg T_c$, the correction to the average spin is proportional to $(T_c/\mu H)^2$. In zero magnetic field, the interaction with conduction electrons also leads to the splitting of doubly degenerate spin impurity states. © *1999 American Institute of Physics.* [S1063-7761(99)00804-5]

1. INTRODUCTION

In the low-temperature and weak magnetic field region, even a weak interaction of magnetic impurities with a degenerate electron gas becomes strong.[1–3] In this region, perturbation theory is violated. Two scenarios are possible in such a situation. First, an assumption can be made that in the low-temperature region, an increase in the magnetic field takes the system out of a strongly coupled state and into the region of applicability of perturbation theory. This nonobvious conjecture was used in Bethe's ansatz method in the problem under consideration. As the result, in a strong magnetic field $\mu H \gg T_c$ (T_c is the Kondo temperature), the correction to the mean spin impurity value has logarithmic behavior,[3]

$$\langle S_z \rangle = \frac{1}{2}\left(1 - \frac{1}{2\ln(\mu H/T_c)}\right).$$

Such spin dependence of the magnetic field value is too slow, and is inconsistent with the experimental data,[4] which yields power-like behavior. The level of spin saturation in the magnetic field in Ref. 4 (Fig. 8) can be reached according to the expression given above only at the magnetic field value $H \approx 50$ T instead of the experimental value of 6 T.

The second scenario is connected with the assumption that an increase only in the magnetic field value does not move the system from a strongly coupled state to a weakly perturbed state. The second conjecture is supported by the fact that the correction to the wave function of a system consisting of magnetic impurity plus degenerate Fermi gas, in some state with low energy, contains corrections of two types obtained with the help of perturbation theory. The norm of one of them decreases in an increasing magnetic field, whereas the norm of the other is divergent in the limit $T \to 0$ for a finite magnetic field. Consideration of the norm of states in the problem involved is very useful, because it contains direct information about the average value of magnetic spin.

Below we consider in detail the second conjecture and confirm it. In the low-temperature region ($T \ll T_c$), the average spin of magnetic impurities is found for an arbitrary value of the external magnetic field. States for both signs of interaction constant are investigated. The strong coupled state arises in both cases, but the magnetic field dependence of the average value of spin is substantially different. The definition of Kondo temperature T_c is also slightly different for different signs of the interaction constant.

2. THE MODEL

We will suppose that the interaction of magnetic impurity with the Fermi sea of electrons has an exchange nature. Then the Hamiltonian $\hat{H}$ of the system under consideration can be taken in the form

$$\hat{H} = \hat{H}_0 + \int d^3r_1 d^3r_2 V(r_1 - r_2) \times \chi_\alpha^+(r_1)\varphi_\beta^+(r_2)\chi_\beta(r_2)\varphi_\alpha(r_1) - \frac{\mu H}{2} \times \int (\varphi_\uparrow^+(r_1)\varphi_\uparrow(r_1) - \varphi_\downarrow^+(r_1)\varphi_\downarrow(r_1)) d^3r_1. \quad (1)$$

In Eq. (1), operators $\varphi_\beta^+, \chi_\alpha^+$ are creation operators of an electron in a localized state on a magnetic impurity and in the continuum spectrum respectively. For simplicity, we consider the case with one unpaired electron in the localized state (spin 1/2). The first term in Eq. (1) describes the degenerate electron gas in some external field that leads to creation of one localized state. The spin interaction of electrons in the continuum spectrum with magnetic field leads only to small renormalization of the magnetic moment of a localized electron, and a small shift in the kinetic energy of electrons with spin up and down in such a way that they have the same value of chemical potential (no gap for transfer of electron

with spin flip over the Fermi level). For this reason we omit this term in Hamiltonian (1). The last term gives the interaction energy of a localized electron with the magnetic field.

Consider now the limiting case as $T\to 0$ and H finite. We search for the lowest-energy eigenfunction $|\psi\rangle$ of Hamiltonian (1) in Fock space in the form

$$
\begin{aligned}
|\psi\rangle = |10;11;11;..\rangle &+ \sum C^{2L-1}_{2K} |01;10;\ ^{2K\ 2L-1}_{10}\rangle \\
&+ \sum C^{2L-1}_{2K-1} |10;\ ^{2K-1\ 2L-1}_{01\ ;\ 10}\rangle + \sum C^{2L}_{2K} |10;10;01\rangle \\
&+ \sum_{K_1<K} C^{2L_1\ 2L-1}_{2K_1\ 2K} \hat{N} |01;10;10;01;\ ^{2K_1\ 2K\ 2L_1\ 2L-1}_{10}\rangle \\
&+ \sum C^{2L_1;2L-1}_{2K_1-1;2K} \hat{N} |10;\ ^{2K_1-1\ 2K\ 2L_1\ 2L-1}_{01\ ;10;01;\ 10}\rangle \\
&+ \sum_{L_1<L} C^{2L_1-1;2L-1}_{2K_1;2K-1} \hat{N} |01;10;\ ^{2K_1\ 2K-1\ 2L_1-1\ 2L-1}_{01\ ;\ 10\ ;\ 10}\rangle \\
&+ \sum_{K_1<K;L_1<L} C^{2L_1-1;2L-1}_{2K_1-1;2K-1} \hat{N} |10;\ ^{2K_1-1\ 2K-1\ 2L_1-1\ 2L-1}_{01\ ;\ 01\ ;\ 10\ ;\ 10}\rangle \\
&+ \sum_{K_1<K;L_1<L} C^{2L_1;2L}_{2K_1;2K} \hat{N} |10;\ ^{2K_1\ 2K\ 2L_1\ 2L}_{10;10;01;01}\rangle + \ldots
\end{aligned}
\tag{2}
$$

In Eq. (2), all single-particle states (solutions of Eq. (1) for one particle) are ordered and numbered. Indexes K, L label states under and over the Fermi surface. Each box has two places. The first one means a state with spin up, and the second with spin down. As an example, the state $^{2K\ 2L}_{10;01}\rangle$ means that the state $2K$ (spin down) under the Fermi surface is empty and the state $2L$ (spin up) over the Fermi surface is filled. The first cell is always reserved for an electron in a localized state. The first term in Eq. (2) gives the ground state of Hamiltonian (1) without interaction ($V(r)=0$). The number of upper (or lower) indexes in $C^{\cdots}_{\cdots}$. gives the number of excited pairs. For P excited pairs, there are $2P+1$ different symbols $C^{\cdots}_{\cdots}$. Operator $\hat{N}$ is the ordering operator, and each rearrangement of two neighboring filled states gives a factor (-1). In Eq. (2) in each box below Fermi surface, only one place can be empty and above the Fermi surface in each box, only one place can be filled.

The equation for the wave function $|\psi\rangle$ is

$$
|\hat{H}\psi\rangle = E|\psi\rangle, \tag{3}
$$

where E is the energy of the state.

Inserting expression (2) for the wave function $|\psi\rangle$ into Eq. (3), we obtain a set of linear equations for the quantities $C^{\cdots}_{\cdots}$. Due to the structure of Hamiltonian (1), each quantity $C^{\cdots}_{\cdots}$ with index P is coupled only with quantities $C^{\cdots}_{\cdots}$ with indexes P, $P\pm 1$. From the first equation of this system, we obtain the energy of the state,

$$
E = E_0 - \mu H/2 - \delta E,
$$

$$
\delta E = \sum \left((I^{2L-1}_{2K-1})^* C^{2L-1}_{2K-1} - (I^{2L-1}_{2K})^* C^{2L-1}_{2K} \right), \tag{4}
$$

where E_0 is the energy of the ground state without interaction. For convenience, we leave the magnetic energy of the localized state out of the correction term δE. The quantities $I^{\cdots}_{\cdots}$ in Eq. (4) are the transition matrix elements. As an example, we have

$$
I^{2L-1}_{2K} = \int d^3r_1 d^3r_2 \chi^*_\uparrow(r_1)\varphi^*_\downarrow(r_2)\varphi_\uparrow(r_1)\chi_\downarrow(r_2)V(r_1-r_2). \tag{5}
$$

The Hamiltonian (1) possesses deep symmetry properties. To see some of these, we will keep indexes on $I^{\cdots}_{\cdots}$ that indicate energy and spin in the initial and final states. The next three equations for the quantities $C^{\cdots}_{\cdots}$ are

$$
\begin{aligned}
&-I^{2L-1}_{2K} + \sum C^{2L-1}_{2K_1} I^{2K_1}_{2K} - \sum C^{2L-1}_{2K_1-1} I^{2K_1-1}_{2K} \\
&\quad - \sum C^{2L_1}_{2K} I^{2L-1}_{2L_1} + (\mu H + \varepsilon_L - \varepsilon_K - \delta E) C^{2L-1}_{2K} \\
&\quad + \sum_{K_1<K} C^{2L_1;2L-1}_{2K_1;2K} I^{2K_1}_{2L_1} - \sum_{K<K_1} C^{2L_1;2L-1}_{2K;2K_1} I^{2K_1}_{2L_1} \\
&\quad - \sum C^{2L_1;2L-1}_{2K_1-1;2K} I^{2K_1-1}_{2L_1} = 0,
\end{aligned}
$$

$$
\begin{aligned}
&I^{2L-1}_{2K-1} - \sum I^{2K_1}_{2K-1} C^{2L-1}_{2K_1} + \sum C^{2L-1}_{2K_1-1} I^{2K_1-1}_{2K-1} \\
&\quad - \sum I^{2L-1}_{2L_1-1} C^{2L_1-1}_{2K-1} + (\varepsilon_L - \varepsilon_K - \delta E) C^{2L-1}_{2K-1} \\
&\quad + \sum_{L_1<L} C^{2L_1-1;2L-1}_{2K_1;2K-1} I^{2K_1}_{2L_1-1} \\
&\quad - \sum_{L<L_1} C^{2L-1;2L_1-1}_{2K_1;2K-1} I^{2K_1}_{2L_1-1} \\
&\quad + \sum_{K<K_1;L_1<L} C^{2L_1-1;2L-1}_{2K-1;2K_1-1} I^{2K_1-1}_{2L_1-1} \\
&\quad - \sum_{K_1<K;L_1<L} C^{2L_1-1;2L-1}_{2K_1-1;2K-1} I^{2K_1-1}_{2L_1-1} \\
&\quad - \sum_{L<L_1;K<K_1} C^{L-1;2L_1-1}_{2K-1;2K_1-1} I^{2K_1-1}_{2L_1-1} \\
&\quad + \sum_{K_1<K;L<L_1} C^{2L-1;2L_1-1}_{2K_1-1;2K-1} I^{2K_1-1}_{2L_1-1} = 0,
\end{aligned}
$$

$$
\begin{aligned}
&- \sum I^{2L}_{2L_1-1} C^{2L_1-1}_{2K} + (\varepsilon_L - \varepsilon_K - \delta E) C^{2L}_{2K} \\
&\quad + \sum_{K<K_1} C^{2L;2L_1-1}_{2K;2K_1} I^{2K_1}_{2L_1-1} - \sum_{K_1<K} C^{2L;2L_1-1}_{2K_1;2K} I^{2K_1}_{2L_1-1} \\
&\quad + \sum C^{2L;2L_1-1}_{2K_1-1;2K} I^{2K_1-1}_{2L_1-1} = 0.
\end{aligned} \tag{6}
$$

In Eqs. (6), the quantities $\varepsilon_{L,K}$ are the energies of single states. As mentioned above, index L means a state above the Fermi level and index K means a state below the Fermi level. The equations for $C^{\cdots}_{\cdots}$ are given in Appendix A. Since the equations for $C^{\cdots}_{\cdots}$ have a special structure, each quantity $C^{\cdots}_{\cdots}$ with index P is coupled only with quantities $C^{\cdots}_{\cdots}$ with indexes, P, $P\pm 1$, it is possible to leave quantities $C^{\cdots}_{\cdots}$ with indexes $P\geqslant 2$ out of Eqs. (6). As the result, we obtain three equations for the quantities C^{2L-1}_{2K}, C^{2L-1}_{2K-1} and C^{2L}_{2K}. They have the following form (see Appendix A):

$$
\begin{aligned}
&-I^{2L-1}_{2K}+\sum C^{2L-1}_{2K_1}I^{2K_1}_{2K}-\sum C^{2L-1}_{2K_1-1}I^{2K_1-1}_{2K}\\
&\quad-\sum C^{2L_1}_{2K}I^{2L-1}_{2L_1}+(\mu H+\varepsilon_L-\varepsilon_K-\delta E\\
&\quad-\Sigma^{(1)}_{(K,L)})C^{2L-1}_{2K}=A_1(C^{2L-1}_{2K};C^{2L-1}_{2K-1};C^{2L}_{2K}),\\
&I^{2L-1}_{2K-1}-\sum C^{2L-1}_{2K_1}I^{2K_1}_{2K-1}+\sum C^{2L-1}_{2K_1-1}I^{2K_1-1}_{2K-1}\\
&\quad-\sum C^{2L_1-1}_{2K-1}I^{2L-1}_{2L_1-1}+(\varepsilon_L-\varepsilon_K-\delta E\\
&\quad-\Sigma_{(K,L)})C^{2L-1}_{2K-1}=A_2(C^{2L-1}_{2K};C^{2L-1}_{2K-1};C^{2L}_{2K}),\\
&-\sum I^{2L}_{2L_1-1}C^{2L_1-1}_{2K}+(\varepsilon_L-\varepsilon_K-\delta E-\Sigma_{(K,L)})C^{2L}_{2K}\\
&\quad=A_3(C^{2L-1}_{2K};C^{2L-1}_{2K-1};C^{2L}_{2K}).
\end{aligned}
\tag{7}
$$

The linear operators $A_{1,2,3}$ do not contain terms proportional to the quantities $C^{2L-1}_{2K},C^{2L-1}_{2K-1},C^{2L}_{2K}$ without integral over one of variable K,L with some function of K,L. These terms form the $\Sigma^{(1)}_{(K,L)},\Sigma_{(K,L)}$ terms in Eq. (7). All off-diagonal elements of such a form are equal to zero. The linear operators $A_{1,2,3}$ also do not contain terms proportional to the convolution of quantities $C^{\cdots}_{\cdots}$ with $I^{\cdots}_{\cdots}$ over one of variable K,L without of denominator with the same variable. In Appendix C, we give the expressions for quantities $\Sigma^{(1)}_{(K,L)},\Sigma_{(K,L)}$ in the fourth order of perturbation theory. Quantities $C^{2L-1}_{2K},C^{2L-1}_{2K-1},C^{2L}_{2K}$ in the third order can be found from equations given in Appendix B. It is easy to check that in the fourth order of perturbation theory,

$$
-\delta E-\Sigma_{(K,L)}\big|_{\varepsilon_K=\varepsilon_L=\varepsilon_F}=0. \tag{8}
$$

This equality holds in all the orders of perturbation theory. Below, we put

$$
-\delta E-\Sigma_{(K,L)}\big|_{\varepsilon_K=\varepsilon_L=\varepsilon_F}=\Delta. \tag{9}
$$

In Eq. (9), $\Delta\equiv\Delta(H)$ is some function of the magnetic field that must be determined from self-consistency. This equation is given below. Very important properties follow from the normalization of states defined by Eqs. (2) and (7). To simplify the investigation of Eqs. (7), we give also the expression for operators $A_{1,2,3}$ in the lowest order of perturbation theory in Appendix D. All statements made above are independent of the exact form of spectrum $\varepsilon_K,\varepsilon_L$ and potential $V(r)$.

3. WAVE FUNCTION OF THE GROUND STATE

The average electron spin $\langle S_z\rangle$ in a bound state at zero temperature can be found by differentiating the energy δE with respect to μH

$$
\langle S_z\rangle=\frac{1}{2}-\frac{\partial\delta E}{\partial\mu H}. \tag{10}
$$

In accordance with quantum mechanical rules, the quantity $\langle S_z\rangle$ in the ground state is also given by an expression containing only norms of the states in expansion (2):

$$
\begin{aligned}
\langle S_z\rangle=\frac{1}{2}\{&1+|C^{2L-1}_{2K-1}|^2+|C^{2L}_{2K}|^2-|C^{2L-1}_{2K}|^2\\
&+|C^{2L_1;2L-1}_{2K_1-1;2K}|^2+|C^{2L_1-1;2L-1}_{2K_1-1;2K-1}|^2+|C^{2L_1;2L}_{2K_1;2K}|^2\\
&-|C^{2L_1;2L-1}_{2K_1;2K}|^2-|C^{2L_1-1;2L-1}_{2K_1;2K-1}|^2+\ldots\}\{1\\
&+|C^{2L-1}_{2K-1}|^2+|C^{2L}_{2K}|^2+|C^{2L-1}_{2K}|^2+|C^{2L_1;2L-1}_{2K_1-1;2K}|^2\\
&+|C^{2L_1-1;2L-1}_{2K_1-1;2K-1}|^2+|C^{2L_1;2L}_{2K_1;2K}|^2+|C^{2L_1;2L-1}_{2K_1;2K}|^2\\
&+|C^{2L_1-1;2L-1}_{2K_1;2K-1}|^2+\ldots\}^{-1}.
\end{aligned}
\tag{11}
$$

Below we use both Eqs. (10) and (11). To solve Eqs. (7) and (9), we consider Δ as a parameter. Then the right-hand sides of Eqs. (7) can be taken into account in perturbation theory. In the leading approximation we obtain

$$
\begin{aligned}
&-I^{2L-1}_{2K}+\sum C^{2L-1}_{2K_1}I^{2K_1}_{2K}-\sum C^{2L-1}_{2K_1-1}I^{2K_1-1}_{2K}\\
&\quad-\sum C^{2L_1}_{2K}I^{2L-1}_{2L_1}+(\mu H+\varepsilon_L-\varepsilon_K+\Delta)C^{2L-1}_{2K}=0,\\
&I^{2L-1}_{2K-1}-\sum C^{2L-1}_{2K_1}I^{2K_1}_{2K-1}+\sum C^{2L-1}_{2K_1-1}I^{2K_1-1}_{2K-1}\\
&\quad-\sum C^{2L_1-1}_{2K-1}I^{2L-1}_{2L_1-1}+(\varepsilon_L-\varepsilon_K+\Delta)C^{2L-1}_{2K-1}=0,\\
&-\sum I^{2L}_{2L_1-1}C^{2L_1-1}_{2K}+(\varepsilon_L-\varepsilon_K+\Delta)C^{2L}_{2K}=0.
\end{aligned}
\tag{12}
$$

Below we make the usual assumptions about the energy-independent value of the density of states near the Fermi surface, and that the characteristic energy in transition matrix elements $I^{\cdots}_{\cdots}$ is also the Fermi energy ε_F. As a result, we can put

$$
\begin{aligned}
&\sum_K I^{\cdots}_{2K}(\ldots)\to g\int_0^{\varepsilon_F}dx(\ldots),\quad \sum_L I^{2L}_{\cdots}\to g\int_0^{A\varepsilon_F}dy(\ldots),\\
&\varepsilon_L-\varepsilon_F=y,\quad \varepsilon_F-\varepsilon_K=x.
\end{aligned}
\tag{13}
$$

In Eqs. (13), g is the dimensionless coupling constant. The potential $V(r)$ in Hamiltonian (1) is in natural units,

hence the smallness of the coupling constant g is connected only to the small radius of bound state.

Due to the energy independence of the transition matrix elements $I^{\cdots}_{\cdots}$, Eqs. (12) can be substantially simplified. To do this, we define new quantities that are convolutions of functions $C^{\cdots}_{\cdots}$ with overlap integral $I^{\cdots}_{\cdots}$ over only one variable, K or L, that is

$$Z_L=\sum_{K_1} I^{2K_1}_{2K} C^{2L-1}_{2K_1}, \quad Z_K=\sum_{L_1} I^{2L}_{2L_1-1} C^{2L_1-1}_{2K},$$

$$Y_L=\sum_{K_1} I^{2K_1-1}_{2K} C^{2L-1}_{2K_1-1}, \quad Y_K=\sum_{L_1} I^{2L-1}_{2L_1-1} C^{2L_1-1}_{2K-1},$$

$$X_L=\sum_{L_1} I^{2L}_{2L_1-1} C^{2L_1}_{2K}, \quad X_K=\sum_{L_1} I^{2L-1}_{2L_1} C^{2L_1}_{2K}. \tag{14}$$

Inserting Eqs. (14) into Eqs. (12), we obtain

$$C^{2L-1}_{2K}=\frac{1}{\mu H+y+x+\Delta}\{I-Z_L+Y_L+X_K\},$$

$$C^{2L-1}_{2K-1}=\frac{1}{y+x+\Delta}\{-I+Z_L-Y_L+Y_K\},$$

$$C^{2L}_{2K}=\frac{1}{y+x+\Delta}\, Z_K, \tag{15}$$

where I is the value of the transition matrix element $I^{\cdots}_{\cdots}$ for states near the Fermi surface. Now from Eqs. (14) and (15) we can obtain a complete set of equations for the quantities $Z_{K,L}, Y_{K,L}, X_{K,L}$ only. In addition, the quantities $X_{K,L}$ are very simply related to $Z_{K,L}, Y_{K,L}$. Eliminating them, we obtain a set of equations for just the quantities $Z_{K,L}, Y_{K,L}$:

$$Z_L\left(1+g\ln\frac{\varepsilon_F}{\mu H+y+\Delta}\right)-Y_L g\ln\frac{\varepsilon_F}{\mu H+y+\Delta}$$
$$=Ig\ln\frac{\varepsilon_F}{\mu H+y+\Delta}+g^2\int_0^{\varepsilon_F}\frac{dx Z_K\ln[A\varepsilon_F/(x+\Delta)]}{\mu H+y+x+\Delta},$$

$$Z_K\left(1-g^2\ln\frac{A\varepsilon_F}{x+\Delta}\ln\frac{A\varepsilon_F}{\mu H+x+\Delta}\right)=Ig\ln\frac{A\varepsilon_F}{\mu H+x+\Delta}$$
$$-g\int_0^{A\varepsilon_F}\frac{dy(Z_L-Y_L)}{\mu H+y+x+\Delta},$$

$$Y_L\left(1+g\ln\frac{\varepsilon_F}{y+\Delta}\right)-Z_L g\ln\frac{\varepsilon_F}{y+\Delta}=-Ig\ln\frac{\varepsilon_F}{y+\Delta}$$
$$+g\int_0^{\varepsilon_F}\frac{dx Y_K}{y+x+\Delta},$$

$$Y_K\left(1-g\ln\frac{A\varepsilon_F}{x+\Delta}\right)=-Ig\ln\frac{A\varepsilon_F}{x+\Delta}$$
$$+g\int_0^{A\varepsilon_F}\frac{dy(Z_L-Y_L)}{y+x+\Delta}. \tag{16}$$

Equations (16) are valid for both signs of the interaction constant g. But their solutions are substantially different for $g<0$ and $g>0$. Consider first the case $g<0$ (attractive interaction in the Kondo problem). In such a case, the quantities Z_L, Y_L are large in comparison with Z_K and Y_K. To obtain this, we introduce a formal definition of ''Kondo'' temperature T_c,

$$|g|\ln\frac{\varepsilon_F}{T_c}=\frac{1}{2}. \tag{17}$$

Now we also put

$$T_L(y)=Z_L-Y_L. \tag{18}$$

Eliminating terms Z_K, Y_K from (16); we obtain one equation the quantity T_L:

$$T_L(y)=\frac{1}{1+g\ln[\varepsilon_F/(y+\Delta)]+g\ln[\varepsilon_F/(\mu H+y+\Delta)]}$$
$$\times\left\{Ig\left(\ln\frac{\varepsilon_F}{y+\Delta}+\ln\frac{\varepsilon_F}{\mu H+y+\Delta}\right)\right.$$
$$+\frac{Ig}{2}\int_0^{\varepsilon_F}dx\left(\frac{1}{\mu H+y+x+\Delta}+\frac{1}{y+x+\Delta}\right)$$
$$\times\left[\frac{g^2\ln[A\varepsilon_F/(x+\Delta)]\ln[A\varepsilon_F/(\mu H+x+\Delta)]}{1-g^2\ln[A\varepsilon_F/(x+\Delta)]\ln[A\varepsilon_F/(\mu H+x+\Delta)]}\right.$$
$$\left.+\frac{g\ln[A\varepsilon_F/(x+\Delta)]}{1-g\ln[A\varepsilon_F/(x+\Delta)]}\right]$$
$$-\frac{g^2}{2}\int_0^{\varepsilon_F}dx\left(\frac{1}{\mu H+y+x+\Delta}+\frac{1}{y+x+\Delta}\right)$$
$$\times\left[\frac{g\ln[A\varepsilon_F/(x+\Delta)]}{1-g^2\ln[A\varepsilon_F/(x+\Delta)]\ln[A\varepsilon_F/(\mu H+x+\Delta)]}\right.$$
$$\times\int_0^{A\varepsilon_F}\frac{dy_1 T_L(y_1)}{\mu H+y_1+x+\Delta}$$
$$\left.\left.+\frac{1}{1-g\ln[A\varepsilon_F/(x+\Delta)]}\int_0^{A\varepsilon_F}\frac{dy_1 T_L(y_1)}{y_1+x+\Delta}\right]\right\}. \tag{19}$$

It can be shown that the last term in Eq. (19) can be omitted, because it is small in the parameter $(|g|\ln(1/|g|))$. We then obtain from Eqs. (17) and (19)

$$T_L(y)=\frac{1}{|g|\ln[(y+\Delta)(\mu H+y+\Delta)/T_c^2]}$$
$$\times\left\{-I-I\int_0^{1/2}dt\left(\frac{t^2}{1-t^2}-\frac{t}{1-t}\right)\right\}. \tag{20}$$

And finally

$$T_L(y)=\frac{-I\beta}{|g|\ln[(y+\Delta)(\mu H+y+\Delta)/T_c^2]},$$

$$\beta=\frac{1}{2}\ln 3+\ln\frac{3}{2}. \tag{21}$$

Inserting Eqs. (18) and (21) into Eqs. (15), we obtain the expressions for coefficients $C_{2K}^{2L-1}, C_{2K-1}^{2L-1}$:

$$C_{2K}^{2L-1}=-\frac{T_L(y)}{\mu H+y+x+\Delta},\quad C_{2K-1}^{2L-1}=\frac{T_L(y)}{y+x+\Delta}. \tag{22}$$

Now we can determine the value of Δ. Equations (10) and (11) should give the same value for average spin $\langle S_z\rangle$. This condition, with the help of Eqs. (4) and (22), gives

$$\beta\int_0^\infty\frac{dy}{(\mu H+y+\Delta)\ln^2[(y+\Delta)(\mu H+y+\Delta)/T_c^2]}\Bigg/ \times\left(1+\frac{\beta^2}{\ln[\Delta(\mu H+\Delta)/T_c^2]}\right) =\frac{\partial\Delta}{\partial\mu H}\frac{1}{\ln[\Delta(\mu H+\Delta)/T_c^2]} +\int_0^\infty\frac{dy}{(\mu H+y+\Delta)\ln^2[(y+\Delta)(\mu H+y+\Delta)/T_c^2]}. \tag{23}$$

We seek a solution of Eq. (23) in the form

$$\Delta(\mu H+\Delta)=T_c^2(1+\gamma),\quad 0<\gamma\ll 1. \tag{24}$$

Terms proportional to γ^{-1} cancel on the right-hand side of Eq. (23). This condition yields

$$\frac{\partial\Delta}{\partial\mu H}+\frac{T_c^2}{(\mu H+\Delta)(\mu H+2\Delta)}=0. \tag{25}$$

The solution of this equation is

$$\Delta(\mu H+\Delta)=T_c^2, \tag{26}$$

$$\Delta=-\frac{\mu H}{2}+\left(\left(\frac{\mu H}{2}\right)^2+T_c^2\right)^{1/2}, \tag{26a}$$

and confirms our conjecture (24) about it. Of course, Eq. (24) has two solutions for Δ. One is given by Eq. (26a) (ground state), and the other is

$$\Delta=-\frac{\mu H}{2}-\left(\left(\frac{\mu H}{2}\right)^2+T_c^2\right)^{1/2}. \tag{26b}$$

Solution (26b) corresponds to the excited state. In the limit $\mu H\gg T_c$, this state transforms to a state with spin orientation along the magnetic field. The excited state is separated from the ground state by a ''gap'' $2((\mu H/2)^2+T_c^2)^{1/2}$. The gap results in the independence of the position of the maximum of impurity heat capacity from the magnetic field in the range $\mu H\ll T_c$ (Schottky anomaly). Such a residual Schottky anomaly is always presented in experiments.[5] In Sec. 4 we will show that renormalization of the term μH in (7) leads to a change from μH in Eq. (27) to $\mu\tilde{H}$ defined by Eq. (43). As a result, we obtain the mean spin $\langle S_z\rangle$ as an implicit function of the magnetic field μH.

An attempt to obtain such an equation at nonzero temperature was made in Ref. 6. But the mean field approximation used there is incorrect for the problem considered.

In Appendix D we show that the right-hand side of (7) leads to renormalization of the coefficients in Eqs. 16, but does not alter the main result of the paper, Eqs. (27) and (43). Of course, renormalization changes Eqs. (17) for the Kondo temperature. The quantity γ can be found only from correction terms to Eqs. (20) and (22). Fortunately, we do not need these correction terms, because in the leading approximation, γ also drops out of Eq. (11) for the spin value. With the help of Eqs. (11), (22), and (24), we obtain

$$\langle S_z\rangle=\frac{\mu H}{2}\int_0^\infty\frac{dy}{(y+\Delta)(y+\mu H+\Delta)(\gamma+y(\mu H+2\Delta)/T_c^2)^2}\Bigg/$$

$$\frac{1}{\gamma}=\frac{\mu H}{4(T_c^2+(\mu H/2)^2)^{1/2}}. \tag{27}$$

Equation (27) is in good agreement with the experimental data of Ref. 4.

4. FERROMAGNETIC CASE ($g>0$)

As mentioned above, Eqs. (16) are valid for both signs of the «interaction» constant g. In the case $g>0$, we can define the characteristic energy of the problem from the relation

$$g\ln\frac{A\varepsilon_F}{T_c}=1 \tag{28}$$

for Kondo temperature T_c. For $g>0$, the quantities Z_K, Y_K, X_K are large in comparison with Z_L, Y_L, X_L. We can eliminate Z_L, Y_L from Eqs. (16). As a result, we have

$$Z_K\left(1-g^2\ln\frac{A\varepsilon_F}{x+\Delta}\ln\frac{A\varepsilon_F}{\mu H+x+\Delta}\right)=Ig\ln\frac{A\varepsilon_F}{\mu H+x+\Delta} -g\int_0^{A\varepsilon_F}\frac{dy}{\mu H+y+x+\Delta} \times\frac{1}{1+g\ln[\varepsilon_F/(y+\Delta)]+g\ln[\varepsilon_F/(\mu H+y+\Delta)]} \times\Bigg[Ig\ln\frac{\varepsilon_F^2}{(y+\Delta)(\mu H+y+\Delta)} +g^2\int_0^{\varepsilon_F}\frac{dx_1 Z_K(x_1)\ln[A\varepsilon_F/(x_1+\Delta)]}{\mu H+y+x_1+\Delta} -g\int_0^{\varepsilon_F}\frac{dx_1 Y_K(x_1)}{y+x_1+\Delta}\Bigg],$$

$$
\begin{aligned}
&Y_K\left(1-g\ln\frac{A\varepsilon_F}{x+\Delta}\right)\\
&\quad=-Ig\ln\frac{A\varepsilon_F}{x+\Delta}+g\int_0^{A\varepsilon_F}\frac{dy}{y+x+\Delta}\\
&\quad\times\frac{1}{1+g\ln[\varepsilon_F/(y+\Delta)]+\ln[\varepsilon_F/(\mu H+y+\Delta)]}\\
&\quad\times\left[Ig\ln\frac{\varepsilon_F^2}{(y+\Delta)(\mu H+y+\Delta)}\right.\\
&\quad+g^2\int_0^{\varepsilon_F}\frac{dx_1 Z_K(x_1)\ln[A\varepsilon_F/(x_1+\Delta)]}{\mu H+y+x_1+\Delta}\\
&\quad\left.-g\int_0^{\varepsilon_F}\frac{dx_1 Y_K(x_1)}{y+x_1+\Delta}\right].
\end{aligned}
\tag{29}
$$

In the range $x\ll\varepsilon_F$, Eqs. (29) yield the following values for Y_K, Z_K:

$$
Y_K=-\frac{ID}{g\ln[(x+\Delta)/T_c]},
$$

$$
Z_K=\frac{ID}{g\ln[(x+\Delta)(\mu H+x+\Delta)/T_c^2]},
\tag{30}
$$

where D is a number of order unity. Inserting Eqs. (30) into Eqs. (15), we obtain

$$
C_{2K}^{2L-1}=\frac{1}{\mu H+y+x+\Delta}\,\frac{ID}{g\ln[(x+\Delta)(\mu H+x+\Delta)/T_c^2]},
$$

$$
C_{2K-1}^{2L-1}=-\frac{1}{y+x+\Delta}\,\frac{ID}{g\ln[(x+\Delta)/T_c]},
$$

$$
C_{2K}^{2L}=\frac{1}{y+x+\Delta}\,\frac{ID}{g\ln[(x+\Delta)(\mu H+x+\Delta)/T_c^2]}.
\tag{31}
$$

In the same way as in the case $g<0$, with the help of Eqs. (10), (11), and (31), we obtain

$$
\begin{aligned}
&\frac{\partial\Delta}{\partial\mu H}\left[\frac{1}{\ln(\Delta/T_c)}+\frac{1}{\ln[\Delta(\mu H+\Delta)/T_c^2]}\right]\\
&\quad=-\left[1-\frac{D^2}{1+D^2\left(\dfrac{1}{\ln(\Delta/T_c)}+\dfrac{1}{\ln[\Delta(\mu H+\Delta)/T_c^2]}\right)}\right]\\
&\quad\times\int_0^{\infty}\frac{dx}{(x+\Delta+\mu H)\ln^2[(x+\Delta)(\mu H+x+\Delta)/T_c^2]}.
\end{aligned}
\tag{32}
$$

The solution of this equation is

$$
\Delta\equiv T_c.
\tag{33}
$$

Relation (33) means that in the leading approximation, the spin value in the magnetic field is saturated,

$$
\langle S_z\rangle=\frac{1}{2}.
\tag{34}
$$

Correction terms to Eq. (34) come only from an energy range ε of order $\varepsilon\propto\varepsilon_F\exp(-1/g^2)$. Note that a similar energy scale also arises in the problem considered by Nozieres and Dominicis.[7] Our conjecture is that in temperature range

$$
T_c^2/\varepsilon_F\ll T\ll T_c,
\tag{35}
$$

the leading correction to the average spin arises from the cutoff of integrals with respect to energy in expression (11) over an energy range of order T. If such an assumption is true, then the average spin in the magnetic field $\mu H\gg T$ is

$$
\begin{aligned}
\langle S_z\rangle&=\frac{1}{2}-\frac{T}{T_c}\\
&\quad\times\int_0^{\infty}\frac{dx}{(x+1+\mu H/T_c)\ln^2[(1+x)(x+1+\mu H/T_c)]}\\
&=\frac{1}{2}-\frac{T}{4T_c}\\
&\quad\times\int_{\ln(1+\mu H/T_c)}^{\infty}\frac{dz}{\overline{[z+(1/2)\ln[1-(\mu H/T_c)e^{-z}]]^2}}.
\end{aligned}
\tag{36}
$$

In the limiting cases of weak ($\mu H\ll T_c$) and strong ($\mu H\gg T_c$) magnetic fields, the average spin is

$$
\langle S_z\rangle=\frac{1}{2}\left(1-\frac{T}{\mu H}\right),\quad T\ll\mu H\ll T_c,
$$

$$
\langle S_z\rangle=\frac{1}{2}\left(1-\frac{T}{2T_c\ln(\mu H/T_c)}\right),\quad \mu H\gg T_c.
\tag{37}
$$

5. SELF-ENERGY TERMS $\Sigma^{(1)}_{(K,L)}, \Sigma_{(K,L)}$ IN PERTURBATION THEORY

As mentioned in the Sec. 1, there are two self-energy terms in the problem under consideration, $\Sigma^{(1)}_{(K,L)}$ and $\Sigma_{(K,L)}$. In second-order perturbation theory, they coincide. They start to differ in third-order in the coupling constant. In third-order perturbation theory, we obtain from Appendix C

$$\Sigma^{(1)}_{(K,L)}-\Sigma_{(K,L)}$$

$$=I^{2K_1}_{2L_1}I^{2L_1}_{2L_2}I^{2L_2}_{2K_1}$$

$$\times\left(\frac{1}{(\mu H+\varepsilon_L+\varepsilon_{L_1}-\varepsilon_K-\varepsilon_{K_1})(\mu H+\varepsilon_L+\varepsilon_{L_2}-\varepsilon_K-\varepsilon_{K_1})}\right.$$

$$\left.-\frac{1}{(\varepsilon_L+\varepsilon_{L_1}-\varepsilon_K-\varepsilon_{K_1})(\varepsilon_L+\varepsilon_{L_2}-\varepsilon_K-\varepsilon_{K_1})}\right). \tag{38}$$

A simple calculation of sums in Eq. (38) leads to

$$(\Sigma^{(1)}_{(K,L)}-\Sigma_{(K,L)})_{\varepsilon_L=\varepsilon_K=\varepsilon_F}=-\mu Hg^3\ln^2\left(\frac{\varepsilon_F}{\varepsilon_c}\right), \tag{39}$$

where ε_c is the cutoff energy. In Appendix C, we obtain the following term in expansion (39) for the self-energy:

$$(\Sigma^{(1)}_{(K,L)}-\Sigma_{(K,L)})_{\varepsilon_K=\varepsilon_L=\varepsilon_F}$$

$$=-\mu Hg^3\ln^2\left(\frac{\varepsilon_F}{\varepsilon_c}\right)+2\mu Hg^4\ln^3\left(\frac{\varepsilon_F}{\varepsilon_c}\right)-\ldots \tag{40}$$

Comparison with the expression for δE obtained in perturbation theory shows that

$$\delta\Sigma=(\Sigma^{(1)}_{(K,L)}-\Sigma_{(K,L)})_{\varepsilon_K=\varepsilon_L=\varepsilon_F}$$

$$=\mu Hg\ln\left(\frac{\varepsilon_F}{\varepsilon_c}\right)\left[-\frac{\partial\delta E}{\partial\mu H}\right]=-\frac{\mu H}{2}\left(-\frac{1}{2}+\langle S_z\rangle\right). \tag{41}$$

To obtain Eq. (41) we used Eqs. (17), (10), and an assumption that $\varepsilon_c\sim T_c$.

Equation (41) means that some corrections should be made in the first of Eqs. (12). Specifically, μH in the first of Eqs. (12) should be corrected by $\delta\Sigma$:

$$\mu H\to\mu H-\delta\Sigma=\mu\tilde{H}. \tag{42}$$

The main result of this correction is a decrease in the initial slope of the magnetic field dependence of the average spin value by 3/4. This phenomenon was probably found in the experimental Ref. 4 (Figs. 8 and 9). The average spin $\langle S_z\rangle$ is given by Eq. (27) with the substitution

$$\mu H\to\mu\tilde{H}=\mu H-\frac{\mu H}{2}\left(\frac{1}{2}-\langle S_z\rangle\right). \tag{43}$$

This equation determines $\langle S_z\rangle$ as an implicit function of μH. From Eqs. (27) and (43), we find that $\langle S_z\rangle$ as a function of μH has an a inflection point at $\mu H/2T_c=0.2426$. Such an inflection point was obtained in Ref. 4.

6. CONCLUSION

Thus, we show that at zero temperature and finite magnetic field $\mu H\ll\varepsilon_F$, a singularity exists in the convolution of amplitudes $C^{2L-1}_{2K_1}$ and $C^{2L-1}_{2K_1-1}$ over energy ε_{K_1} with amplitude $I^{2K_1}_{2K}$. As a result, in the high magnetic field region, $\mu H\gg T_c$, the correction to the spin impurity value is proportional to $(T_c/\mu H)^2$ instead of $1/\ln(\mu H/T_c)$, as predicted in Refs. 1–3. We also find that renormalization of the magnetic field discussed in Sec. 4 leads to an inflection point in the dependence of spin impurity on the magnetic field. The initial slope is a function of z, which enters into the definition of the Kondo temperature (see Appendix D). Our consideration shows that the interaction of the spin of an impurity with an electron gas does not lead to the ap pearance of the localized state, as assumed in Refs. 8–10. The Kondo temperature T_c is given by Eq. (D7), where z is the root of the equation

$$f(z)=0. \tag{44}$$

We find here three terms in the expansion of f in Taylor series [Eq. (D8)]. This equation was also studied in Refs. 8 and 11. Our result for the first two terms in Eq. (44) coincide with the result of Ref. 11, because this is also the result of parquet approximation. But, our consideration [Eq. (44)] is conceptually closer to the Ref. 8. The difference even in the second term is probably related to the assumption of Ref. 8 that in the problem under consideration there is a localized state with spin 1/2.

In fact, such a localized state does not exist. Without interaction there are two states associated with impurity spin 1/2. In zero magnetic field, these two states are degenerate. Interaction removes such a degeneracy and the splitting energy is $2T_c$. Of course, interaction does not change the number of states, as in our consideration, and is not fulfilled in Ref. 8. Note also that the driving term is missing in Refs. 8–10.

Nevertheless, the average value of spin of impurity $\langle S_z\rangle$ as a function of magnetic field found in Refs. 9 and 10 coincides with our result except for the effect of renormalization of the magnetic field (Sec. 4) and the expression for the Kondo temperature.

The authors thank Prof. P. Fulde and Prof. A. I. Larkin for helpful discussions. We thank Prof. P. Fulde for hospitality at the Max-Planck-Institute for Complex Systems (Dresden). The research of Yu.N.O. was supported by CRDF (Grant No RP1-194). The research of A.M.D. is supported by INTAS and the Russian Fund for Fundamental Research (Grant No 95-553).

APPENDIX A

The wave function of a system consisting of one localized electron plus degenerate electron gas can be taken in the form

$$
\begin{aligned}
|\psi\rangle = |10;11;11...\rangle &+ \sum C^{2L-1}_{2K} |01;10;\overset{2K\ 2L-1}{10\ \ 10}\rangle + \sum C^{2L-1}_{2K-1} |10;\overset{2K-1\ 2L-1}{01\ \ 10}\rangle + \sum C^{2L}_{2K} |10;10;01\rangle \\
&+ \sum_{K_1<K} C^{2L_1;2L-1}_{2K_1;2K} \hat{N}|01;\overset{2K_1\ 2K\ 2L_1\ 2L-1}{10;10;01;\ 10}\rangle + \sum C^{2L_1;2L-1}_{2K_1-1;2K} \hat{N}|10;\overset{2K_1-1\ 2K\ 2L_1\ 2L-1}{01\ \ ;10;01;\ 10}\rangle \\
&+ \sum_{L_1<L} C^{2L_1-1;2L-1}_{2K_1;2K-1} \hat{N}|01;\overset{2K_1\ 2K-1\ 2L_1-1\ 2L-1}{10;\ 01\ ;\ 10\ ;\ 10}\rangle + \sum_{K_1<K;L_1<L} C^{2L_1-1;2L-1}_{2K_1-1;2K-1} \hat{N}|10;\overset{2K_1-1\ 2K-1\ 2L_1-1\ 2L-1}{01\ ;\ 01\ ;\ 10\ ;\ 10}\rangle \\
&+ \sum_{K_1<K;L_1<L} C^{2L_1;2L}_{2K_1;2K}|10;\overset{2K_1\ 2K\ 2L_1\ 2L}{10;10;01;01}\rangle + \sum_{K_1<K_1<K;L_2<L_1} C^{2L_1;2L_1;2L-1}_{2K_2;2K_1;2K} \hat{N}|01;\overset{2K_2\ 2K_1\ 2K\ 2L_2 2L_1\ 2L-1}{10;\ 10;10;\ 0101;\ 10}\rangle \\
&+ \sum_{K_1<K;L_2<L_1} C^{2L_2;2L_1;2L-1}_{2K_2-1;2K_1;2K} \hat{N}|10;\overset{2K_2-1\ 2K_1\ 2K\ 2L_2 2L_1\ 2L-1}{01\ \ ;10;10;\ 0101;\ 10}\rangle \\
&+ \sum_{K_2<K;L_1<L} C^{2L_2;2L_1-1;2L-1}_{2K_2-1;2K_1;2K-1} \hat{N}|10;\overset{2K_2-1\ 2K_1\ 2K-1\ 2L_2 2L_1-1\ 2L-1}{01\ \ ;10;\ 01\ ;\ 0101\ ;\ 10}\rangle \\
&+ \sum_{K_2<K_1;L_1<L} C^{2L_2;2L_1-1;2L-1}_{2K_2;2K_1;2K-1} \hat{N}|01;\overset{2K_2\ 2K_1\ 2K-1\ 2L_2 2L_1-1\ 2L-1}{10;10;\ 01\ ;\ 01\ \ 10\ ;\ 10}\rangle \\
&+ \sum_{K_1<K;L_2<L_1<L} C^{2L_2-1;2L_1-1;2L-1}_{2K_1;2K_1-1;2K-1} \hat{N}|01;\overset{2K_2\ 2K_1-1\ 2K-1\ 2L_2-12L_1-1\ 2L-1}{10;\ 01\ ;\ 01\ ;\ \ 10\ 10\ \ ;\ 10}\rangle \\
&+ \sum_{K_2<K_1<K;L_2<L_1<L} C^{2L_2-1;2L_1-1;2L-1}_{2K_2-1;2K_1-1;2K-1} \hat{N}|10;\overset{2K_2-1\ 2K_1-1\ 2K-1\ 2L_2-12L_1-1\ 2L-1}{01\ \ ;\ 01\ ;\ 01\ ;\ \ 10\ 10\ \ ;\ 10}\rangle \\
&+ \sum_{K_2<K_1<K;L_2<L_1<L} C^{2L_2;2L_1;2L}_{2K_2;2K_1;2K} \hat{N}|10;\overset{2K_2-1\ 2K_1-1\ 2K-1\ 2L_2 2L_1\ 2L}{10\ \ ;\ 10\ \ ;\ 10\ ;0101;01}\rangle + \ldots
\end{aligned}
\tag{A1}
$$

The notations here are the same as in the text. As we note above, there are $(2P+1)$ different symbols $C^{\cdots}_{\cdots}$ of order P. Inserting Eq. (A1) into Eq. (3) for the wave function, some simple but tedions calculations yield a set of equations for the coefficients $C^{\cdots}_{\cdots}$. The five equations for the $C^{\cdots}_{\cdots}$ are

$$
\begin{aligned}
&C^{2L-1}_{2K} I^{2L_1}_{2K_1} - C^{2L-1}_{2K_1} I^{2L_1}_{2K} - C^{2L_1}_{2K} I^{2L-1}_{2K_1} + C^{2L_1}_{2K_1} I^{2L-1}_{2K} + (\mu H + \varepsilon_L + \varepsilon_{L_1} - \varepsilon_K - \varepsilon_{K_1} - \delta E) C^{2L_1;2L-1}_{2K_1;2K} + \left(\sum_{K_2<K} C^{2L_1;2L-1}_{2K_1;2K} I^{2K_2}_{2K_1} \right. \\
&\quad \left. - \sum_{K<K_2} C^{2L_1;2L-1}_{2K;2K_2} I^{2K_2}_{2K_1} \right) + \left(\sum_{K_1<K_2} C^{2L_1;2L-1}_{2K_1;2K_2} I^{2K_2}_{2K} - \sum_{K_2<K_1} C^{2L_1;2L-1}_{2K_2;2K_1} I^{2K_2}_{2K} \right) - \sum C^{2L_2;2L-1}_{2K_1;2K} I^{2L_1}_{2L_2} \\
&\quad + \left(\sum_{L_2<L_1} C^{2L_2;2L_1}_{2K_1;2K} I^{2L-1}_{2L_2} - \sum_{L_1<L_2} C^{2L_1;2L_2}_{2K_1;2K} I^{2L-1}_{2L_2} \right) - \left(\sum C^{2L_1;2L-1}_{2K_2-1;2K} I^{2K_2-1}_{2K_1} - \sum C^{2L_1;2L-1}_{2K_2-1;2K_1} I^{2K_2-1}_{2K} \right) \\
&\quad - \sum_{K_1<K<K_2;L_2<L_1} C^{2L_2;2L_1;2L-1}_{2K_1;2K;2K_2} I^{2K_2}_{2L_2} + \sum_{K_1<K_2<K;L_2<L_1} C^{2L_2;2L_1;2L-1}_{2K_1;2K_2;2K} I^{2K_2}_{2L_2} - \sum_{K_2<K_1;L_2<L_1} C^{2L_2;2L_1;2L-1}_{2K_2;2K_1;2K} I^{2K_1}_{2L_2} \\
&\quad + \sum_{K_1<K<K_2;L_1<L_2} C^{2L_1;2L_2;2L-1}_{2K_1;2K;2K_2} I^{2K_2}_{2L_2} - \sum_{K_1<K_2<K;L_1<L_2} C^{2L_1;2L_2;2L-1}_{2K_1;2K_2;2K} I^{2K_2}_{2L_2} + \sum_{K_2<K_1<K;L_1<L_2} C^{2L_1;2L_2;2L-1}_{2K_2;2K_1;2K} I^{2K_2}_{2L_2} \\
&\quad + \sum_{L_2<L_1;K_1<K} C^{2L_2;2L_1;2L-1}_{2K_2-1;2K_1;2K} I^{2K_2-1}_{2L_2} - \sum_{L_1<L_2;K_1<K} C^{2L_1;2L_2;2L-1}_{2K_2-1;2K_1;2K} I^{2K_2-1}_{2L_2} = 0,
\end{aligned}
$$

$$
\begin{aligned}
&-I^{2L_1}_{2K_1-1}C^{2L-1}_{2K}+C^{2L_1}_{2K}I^{2L-1}_{2K_1-1}-\sum_{K_2<K}C^{2L_1;2L-1}_{2K_2;2K}I^{2K_2}_{2K_1-1}+\sum_{K<K_2}C^{2L_1;2L-1}_{2K;2K_2}I^{2K_2}_{2K_1-1}+(\varepsilon_L+\varepsilon_{L_1}-\varepsilon_K-\varepsilon_{K_1}-\delta E)C^{2L_1;2L-1}_{2K_1-1;2K}\\
&\quad+\sum C^{2L_1;2L-1}_{2K_2-1;2K}I^{2K_2-1}_{2K_1-1}-\sum C^{2L_1;2L_2-1}_{2K_1-1;2K}I^{2L-1}_{2L_2-1}+\sum_{L_2<L}C^{2L_2-1;2L-1}_{2K;2K_1-1}I^{2L_1}_{2L_2-1}-\sum_{L<L_2}C^{2L-1;2L_2-1}_{2K;2K_1-1}I^{2L_1}_{2L_2-1}\\
&\quad+\sum_{L_2<L;K<K_2}C^{2L_1;2L_2-1;2L-1}_{2K;2K_2;2K_1-1}I^{2K_2}_{2L_2-1}-\sum_{L_2<L;K_2<K}C^{2L_1;2L_2-1;2L-1}_{2K_2;2K;2K_1-1}I^{2K_2}_{2L_2-1}-\sum C^{2L_1;2L-1;2L_2-1}_{2K;2K_2;2K_1-1}I^{2K_2}_{2L_2-1}\\
&\quad+\sum_{L<L_2;K_2<K}C^{2L_1;2L-1;2L_2-1}_{2K_2;2K;2K_1-1}I^{2K_2}_{2L_2-1}-\sum_{L_2<L;K_1<K_2}C^{2L_1;2L_2-1;2L-1}_{2K_1-1;2K;2K_2-1}I^{2K_2-1}_{2L_2-1}+\sum_{L_2<L;K_2<K_1}C^{2L_1;2L_2-1;2L-1}_{2K_2-1;2K;2K_1-1}I^{2K_2-1}_{2L_2-1}\\
&\quad+\sum_{L<L_2;K_1<K_2}C^{2L_1;2L-1;2L_2-1}_{2K_1-1;2K;2K_2-1}I^{2K_2-1}_{2L_2-1}-\sum_{L<L_2;K_2<K_1}C^{2L_1;2L-1;2L_2-1}_{2K_2-1;2K;2K_1-1}I^{2K_2-1}_{2L_2-1}=0,\\
&C^{2L-1}_{2K-1}I^{2L_1-1}_{2K}-C^{2L_1-1}_{2K-1}I^{2L-1}_{2K_1}+\sum C^{2L_2;2L-1}_{2K-1;2K_1}I^{2L_1-1}_{2L_2}-\sum C^{2L_2;2L_1-1}_{2K-1;2K_1}I^{2L-1}_{2L_2}+(\mu H+\varepsilon_L+\varepsilon_{L_1}-\varepsilon_K-\varepsilon_{K_1}\\
&-\delta E)C^{2L_1-1;2L-1}_{2K_1;2K-1}+\sum C^{2L_1-1;2L-1}_{2K_2;2K-1}I^{2K_2}_{2K_1}+\sum_{K<K_2}C^{2L_1-1;2L-1}_{2K-1;2K_2-1}I^{2K_2-1}_{2K_1}-\sum_{K_2<K}C^{2L_1-1;2L-1}_{2K_2-1;2K-1}I^{2K_2-1}_{2K_1}\\
&+\sum_{K_1<K_2}C^{2L_2;2L_1-1;2L-1}_{2K_1;2K_2;2K-1}I^{2K_2}_{2L_2}-\sum_{K_2<K_1}C^{2L_2;2L_1-1;2L-1}_{2K_2;2K_1;2K-1}I^{2K_2}_{2L_2}-\sum_{K<K_2}C^{2L_2;2L_1-1;2L-1}_{2K-1;2K_1;2K_1-1}I^{2K_2-1}_{2L_2}\\
&+\sum_{K_2<K}C^{2L_2;2L_1-1;2L-1}_{2K_2-1;2K_1;2K-1}I^{2K_2-1}_{2L_2}=0,\\
&(\varepsilon_L+\varepsilon_{L_1}-\varepsilon_K-\varepsilon_{K_1}-\delta E)C^{2L_1;2L}_{2K_1;2K}+\sum C^{2L;2L_2-1}_{2K_1;2K}I^{2L_1}_{2L_2-1}-\sum C^{2L_1;2L_2-1}_{2K_1;2K}I^{2L}_{2L_2-1}-\sum_{K_1<K<K_2}C^{2L_1;2L;2L_2-1}_{2K_1;2K;2K_2}I^{2K_2}_{2L_2-1}\\
&\quad+\sum_{K_1<K_2<K}C^{2L_1;2L;2L_2-1}_{2K_1;2K_2;2K}I^{2K_2}_{2L_2-1}-\sum_{K_2<K_1<K}C^{2L_1;2L;2L_2-1}_{2K_2;2K_1;2K}I^{2K_2}_{2L_2-1}+\sum C^{2L_1;2L;2L_2-1}_{2K_2-1;2K_1;2K}I^{2K_2-1}_{2L_2-1}=0,\\
&-I^{2L_1-1}_{2K_1-1}C^{2L-1}_{2K-1}+C^{2L-1}_{2K_1-1}I^{2L_1-1}_{2K-1}+C^{2L_1-1}_{2K-1}I^{2L-1}_{2K_1-1}-C^{2L_1-1}_{2K_1-1}I^{2L-1}_{2K-1}+(\varepsilon_L+\varepsilon_{L_1}-\varepsilon_K-\varepsilon_{K_1}-\delta E)C^{2L_1-1;2L-1}_{2K_1-1;2K-1}\\
&\quad-\left(\sum C^{2L_1-1;2L-1}_{2K_2;2K-1}I^{2K_2}_{2K_1-1}-\sum C^{2L_1-1;2L-1}_{2K_2;2K_1-1}I^{2K_2}_{2K-1}\right)+\left(\sum_{K_2<K}C^{2L_1-1;2L-1}_{2K_2-1;2K-1}I^{2K_2-1}_{2K_1-1}-\sum_{K_2<K_1}C^{2L_1-1;2L-1}_{2K_2-1;2K_1-1}I^{2K_2-1}_{2K-1}\right)\\
&\quad-\left(\sum_{K<K_1}C^{2L_1-1;2L-1}_{2K-1;2K_2-1}I^{2K_2-1}_{2K_1-1}-\sum_{K_1<K_2}C^{2L_1-1;2L-1}_{2K_1-1;2K_2-1}I^{2K_2-1}_{2K-1}\right)-\left(\sum_{L_2<L}C^{2L_2-1;2L-1}_{2K_1-1;2K-1}I^{2L_1-1}_{2L_2-1}\right.\\
&\quad\left.-\sum_{L_2<L_1}C^{2L_2-1;2L_1-1}_{2K_1-1;2K-1}I^{2L-1}_{2L_2-1}\right)+\left(\sum_{L<L_2}C^{2L-1;2L_2-1}_{2K_1-1;2K-1}I^{2L_1-1}_{2L_2-1}-\sum_{L_1<L_2}C^{2L_1-1;2L_2-1}_{2K_1-1;2K-1}I^{2L-1}_{2L_2-1}\right)\\
&\quad-\sum_{L_2<L_1<L}C^{2L_2-1;2L_1-1;2L-1}_{2K_2;2K_1-1;2K-1}I^{2K_2}_{2L_2-1}+\sum_{L_1<L_2<L}C^{2L_1-1;2L_2-1;2L-1}_{2K_2;2K_1-1;2K-1}I^{2K_2}_{2L_2-1}-\sum_{L_1<L<L_2}C^{2L_1-1;2L-1;2L_2-1}_{2K_1;2K_1-1;2K-1}I^{2K_1}_{2L_2-1}\\
&\quad+\sum_{L_2<L_1<L;K<K_2}C^{2L_2-1;2L_1-1;2L-1}_{2K_1-1;2K-1;2K_2-1}I^{2K_2-1}_{2L_2-1}-\sum_{K_1<K_2<K;L_2<L_1}C^{2L_2-1;2L_1-1;2L-1}_{2K_1-1;2K_2-1;2K-1}I^{2K_2-1}_{2L_2-1}\\
&\quad+\sum_{K_2<K_1;L_2<L_1}C^{2L_2-1;2L_1-1;2L-1}_{2K_2-1;2K_1-1;2K-1}I^{2K_2-1}_{2L_2-1}-\sum_{K<K_2;L_1<L_2<L}C^{2L_1-1;2L_2-1;2L-1}_{2K_1-1;2K-1;2K_2-1}I^{2K_2-1}_{2L_2-1}\\
&\quad+\sum_{K_1<K_2<K;L_1<L_2<L}C^{2L_1-1;2L_2-1;2L-1}_{2K_1-1;2K_2-1;2K-1}I^{2K_2-1}_{2L_2-1}-\sum_{K_2<K_1;L_1<L_2<L}C^{2L_1-1;2L_2-1;2L-1}_{2K_2-1;2K_1-1;2K-1}I^{2K_2-1}_{2L_2-1}\\
&\quad+\sum_{K_1<K<K_2;L<L_2}C^{2L_1-1;2L-1;2L_2-1}_{2K_1-1;2K-1;2K_2-1}I^{2K_2-1}_{2L_2-1}-\sum_{L<L_2;K_1<K_2<K}C^{2L_1-1;2L-1;2L_2-1}_{2K_1-1;2K_2-1;2K-1}I^{2K_2-1}_{2L_2-1}\\
&\quad+\sum_{K_2<K_1;L<L_2}C^{2L_1-1;2L-1;2L_2-1}_{2K_2-1;2K_1-1;2K-1}I^{2K_2-1}_{2L_2-1}=0.
\end{aligned}
\tag{2}
$$

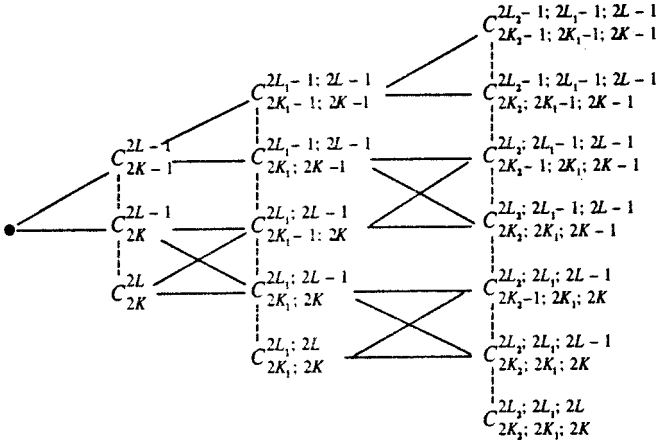

FIG. 1. Relation between various terms $C^{\cdots}_{\cdots}$. Dashed lines represent scattering process.

Equations (A2) are exact.

APPENDIX B

Our purpose is to obtain an expression for the self-energy terms $\Sigma^{(1)}_{(K,L)}$ and $\Sigma_{(K,L)}$ in fourth-order perturbation theory. To do this we should obtain equations on the quantities $C^{\cdots}_{\cdots}$ in the «leading» approximation. Really, we need only six equations in the six quantities entering into Eqs. (A2). The required system can be obtained from Eqs. (3) and (A1). These equations are

$$
\begin{aligned}
&(\mu H+\varepsilon_L+\varepsilon_{L_1}+\varepsilon_{L_2}-\varepsilon_K-\varepsilon_{K_1}-\varepsilon_{K_2})C^{2L_2;2L_1;2L-1}_{2K_2;2K_1;2K}\\
&-\{C^{2L_1;2L-1}_{2K_1;2K}I^{2L_2}_{2K_2}-C^{2L_1;2L-1}_{2K_2;2K}I^{2L_2}_{2K_1}+C^{2L_2;2L-1}_{2K_2;2K_1}I^{2L_2}_{2K}\\
&-C^{2L_2;2L-1}_{2K_1;2K}I^{2L_1}_{2K_2}+C^{2L_2;2L-1}_{2K_2;2K}I^{2L_1}_{2K_1}-C^{2L_2;2L-1}_{2K_2;2K_1}I^{2L_1}_{2K}\}\\
&-\{C^{2L_2;2L_1}_{2K_1;2K}I^{2L-1}_{2K_2}-C^{2L_2;2L_1}_{2K_2;2K}I^{2L-1}_{2K_1}+C^{2L_2;2L_1}_{2K_2;2K_1}I^{2L-1}_{2K}\}=0,\\
&(\varepsilon_L+\varepsilon_{L_1}+\varepsilon_{L_2}-\varepsilon_K-\varepsilon_{K_1}-\varepsilon_{K_2})C^{2L_2;2L_1;2L-1}_{2K_2-1;2K_1;2K}\\
&\quad+C^{2L_2;2L_1}_{2K_1;2K}I^{2L-1}_{2K_2-1}+\{C^{2L_1;2L-1}_{2K_1;2K}I^{2L_2}_{2K_2-1}\\
&\quad-C^{2L_2;2L-1}_{2K_1;2K}I^{2L-1}_{2K_2-1}\}=0.\\
&(\mu H+\varepsilon_L+\varepsilon_{L_1}+\varepsilon_{L_2}-\varepsilon_K-\varepsilon_{K_1}-\varepsilon_{K_2})C^{2L_2;2L_1-1;2L-1}_{2K_2;2K_1;2K-1}\\
&-\{C^{2L_2;2L-1}_{2K-1;2K_1}I^{2L_1-1}_{2K_2}-C^{2L_2;2L-1}_{2K-1;2K_2}I^{2L_1-1}_{2K_1}\\
&-C^{2L_2;2L_1-1}_{2K-1;2K_1}I^{2L-1}_{2K_2}+C^{2L_2;2L_1-1}_{2K-1;2K_2}I^{2L-1}_{2K_1}\}\\
&-\{C^{2L_1-1;2L-1}_{2K_1;2K-1}I^{2L_2}_{2K_2}-C^{2L_1-1;2L-1}_{2K_2;2K-1}I^{2L_2}_{2K_1}\}=0,\\
&(\varepsilon_L+\varepsilon_{L_1}+\varepsilon_{L_2}-\varepsilon_K-\varepsilon_{K_1}-\varepsilon_{K_2})C^{2L_2;2L_1-1;2L-1}_{2K_2-1;2K_1;2K-1}\\
&+\{C^{2L_2;2L-1}_{2K-1;2K_1}I^{2L_1-1}_{2K_2-1}-C^{2L_2;2L-1}_{2K_2-1;2K_1}I^{2L_1-1}_{2K-1}\\
&-C^{2L_2;2L_1-1}_{2K-1;2K_1}I^{2L-1}_{2K_2-1}+C^{2L_2;2L_1-1}_{2K_2-1;2K_1}I^{2L-1}_{2K-1}\}\\
&+\{C^{2L_1-1;2L-1}_{2K_1;2K-1}I^{2L_2}_{2K_2-1}-C^{2L_1-1;2L-1}_{2K_1;2K_2-1}I^{2L_2}_{2K-1}\}=0,\\
&(\mu H+\varepsilon_L+\varepsilon_{L_1}+\varepsilon_{L_2}-\varepsilon_K-\varepsilon_{K_1}\\
&\quad-\varepsilon_{K_2})C^{2L_2-1;2L_1-1;2L-1}_{2K_2;2K_1-1;2K-1}-\{C^{2L_1-1;2L-1}_{2K_1-1;2K-1}I^{2L_2-1}_{2K_2}\\
&\quad-C^{2L_2-1;2L-1}_{2K_1-1;2K-1}I^{2L_1-1}_{2K_2}+C^{2L_2-1;2L_1-1}_{2K_1-1;2K-1}I^{2L-1}_{2K_2}\}=0,\\
&(\varepsilon_L+\varepsilon_{L_1}+\varepsilon_{L_2}-\varepsilon_K-\varepsilon_{K_1}-\varepsilon_{K_2})C^{2L_2-1;2L_1-1;2L-1}_{2K_2-1;2K_1-1;2K-1}\\
&+\{C^{2L_1-1;2L-1}_{2K_1-1;2K-1}I^{2L_2-1}_{2K_2-1}-C^{2L_1-1;2L-1}_{2K_2-1;2K-1}I^{2L_2-1}_{2K_1-1}\\
&+C^{2L_1-1;2L-1}_{2K_2-1;2K_1-1}I^{2L_2-1}_{2K-1}-C^{2L_2-1;2L-1}_{2K_1-1;2K-1}I^{2L_2-1}_{2K_2-1}\\
&+C^{2L_2-1;2L-1}_{2K_2-1;2K-1}I^{2L_1-1}_{2K_1-1}-C^{2L_2-1;2L-1}_{2K_1-1;2K_1-1}I^{2L_1-1}_{2K-1}\\
&+C^{2L_2-1;2L_1-1}_{2K_1-1;2K-1}I^{2L-1}_{2K_2-1}-C^{2L_2-1;2L_1-1}_{2K_2-1;2K-1}I^{2L-1}_{2K_1-1}\\
&+C^{2L_2-1;2L_1-1}_{2K_2-1;2K_1-1}I^{2L-1}_{2K-1}\}=0. \qquad \text{(B1)}
\end{aligned}
$$

Equations (B1) can easily be supplemented by scattering terms, and Eqs. (7), (A1), and (B1) will still form a complete set. The structure of interaction Hamiltonian (1) is such that scattering leads to connection of the given term only with itself and with two (or one) neighboring terms. These terms can be obtained from the given one by a change of parity of one of the upper or lower indexes. The relationships among the various terms $C^{\cdots}_{\cdots}$ are presented in Fig. 1.

APPENDIX C

We are now able to obtain the self-energy parts $\Sigma^{(1)}_{(K,L)}$ and $\Sigma_{(K,L)}$ in fourth-order perturbation theory. Straightforward elimination of terms in $C^{\cdots}_{\cdots}$ with $P\geqslant 2$ from Eqs. (6) using Eqs. (A2) and (B1) gives

$$\Sigma^{(1)}_{(K,L)}=\frac{I^{2K_1}_{2L_1}}{\mu H+\varepsilon_4(L,L_1,K,K_1)-|I^{2L_2}_{2K_2-1}|^2/\varepsilon_6-|I^{2L_2}_{2K_2}|^2/(\mu H+\varepsilon_6)-\delta E}$$

$$\times\left\{I^{2L_1}_{2K_1}-\frac{I^{2K_2}_{2K_1}}{\mu H+\varepsilon_4(L,L_1,K,K_1)}\left(I^{2L_1}_{2K_1}-\frac{I^{2K_3}_{2K_2}I^{2L_1}_{2K_3}}{\mu H+\varepsilon_4(L,L_1,K,K_3)}+\frac{I^{2L_1}_{2L_2}I^{2L_2}_{2K_2}}{\mu H+\varepsilon_4(L,L_2,K,K_2)}\right.\right.$$

$$\left.-\frac{I^{2K_3-1}_{2K_2}I^{2L_1}_{2K_3-1}}{\varepsilon_4(L,L_1,K,K_3)}\right)-\frac{I^{2L_1}_{2L_2}}{\mu H+\varepsilon_4(L,L_2,K,K_1)}\left(-I^{2L_2}_{2K_1}+\frac{I^{2K_2}_{2K_1}I^{2L_2}_{2K_2}}{\mu H+\varepsilon_4(L,L_2,K,K_2)}-\frac{I^{2L_2}_{2L_3}I^{2L_3}_{2K_1}}{\mu H+\varepsilon_4(L,L_3,K,K_1)}\right.$$

$$\left.+\frac{I^{2K_2-1}_{2K_1}I^{2L_2}_{2K_2-1}}{\varepsilon_4(L,L_2,K,K_2)}\right)-\frac{I^{2K_2-1}_{2K_1}}{\varepsilon_4(L,L_1,K,K_2)}\left(I^{2L_1}_{2K_2-1}-\frac{I^{2K_1}_{2K_2-1}I^{2L_1}_{2K_1}}{\mu H+\varepsilon_4(L,L_1,K_3,K)}-\frac{I^{2K_3-1}_{2K_2-1}I^{2L_1}_{2K_3-1}}{\varepsilon_4(L,L_1,K,K_3)}\right)$$

$$-\frac{I^{2K_2}_{2L_2}}{\mu H+\varepsilon_6}\left(\frac{I^{2L_1}_{2K_2}I^{2L_2}_{2K_1}}{\mu H+\varepsilon_4(L,L_2,K,K_1)}+\frac{I^{2L_1}_{2K_1}I^{2L_1}_{2K_2}}{\mu H+\varepsilon_4(L_1,L,K,K_2)}-\frac{I^{2L_1}_{2K_1}I^{2L_2}_{2K_2}}{\mu H+\varepsilon_4(L,L_2,K,K_2)}\right)$$

$$\left.\left.-\frac{I^{2K_2-1}_{2L_2}I^{2L_1}_{2K_2-1}I^{2L_2}_{2K_1}}{\varepsilon_6(\mu H+\varepsilon_6(L,L_2,K,K_1))}\right\}+\frac{I^{2K_1-1}_{2L_1}}{\varepsilon_4(L,L_1,K,K_1)-|I^{2K_2}_{2L_2-1}|^2/(\mu H+\varepsilon_6)-|I^{2K_2-1}_{2L_2-1}|^2/\varepsilon_6-\delta E}\right.$$

$$\times\left\{I^{2L_1}_{2K_1-1}-\frac{I^{2K_2}_{2K_1-1}}{\mu H+\varepsilon_4(L,L_1,K,K_2)}\left(I^{2L_1}_{2K_2}-\frac{I^{2K_3}_{2K_2}I^{2L_1}_{2K_3}}{\mu H+\varepsilon_4(L,L_1,K,K_3)}+\frac{I^{2L_1}_{2L_2}I^{2L_2}_{2K_2}}{\mu H+\varepsilon_4(L,L_2,K,K_2)}-\frac{I^{2K_3-1}_{2K_3}I^{2L_1}_{2K_3-1}}{\varepsilon_4(L,L_1,K,K_3)}\right)\right.$$

$$-\frac{I^{2K_2-1}_{2K_1-1}}{\varepsilon_4(L,L_1,K,K_2)}\left(I^{2L_1}_{2K_2-1}-\frac{I^{2K_3}_{2K_2-1}I^{2L_1}_{2K_3}}{\mu H+\varepsilon_4(L,L_1,K,K_3)}-\frac{I^{2K_3-1}_{2K_2-1}I^{2L_1}_{2K_3-1}}{\varepsilon_4(L,L_1,K,K_3)}\right)$$

$$\left.+\frac{I^{2L-1}_{2L_2-1}I^{2L_3-1}_{2L_3}I^{2L_3}_{2K_1-1}}{(\mu H+\varepsilon_4(L,L_2K,K_1))\varepsilon_4(L,L_3,K,K_1)}-\frac{I^{2K_2-1}_{2L_2-1}I^{2L_2-1}_{2K_1-1}I^{2L_1}_{2K_2-1}}{\varepsilon_6\varepsilon_4(L,L_1,K,K_2)}\right\},\tag{C1}$$

$$\Sigma_{(K,L)}=\frac{I^{2K_1}_{2L_1-1}}{\mu H+\varepsilon_4(L,L_1,K,K_1)-\delta E-|I^{2L_2}_{2K_2-1}|^2/\varepsilon_6-|I^{2L_2}_{2K_2}|^2/(\mu H+\varepsilon_6)}\left\{I^{2L_1-1}_{2K_1}-\frac{I^{2K_2}_{2K_1}}{\mu H+\varepsilon_4(L,L_1,K,K_2)}\right.$$

$$\times\left(I^{2L_1-1}_{2K_2}-\frac{I^{2K_3}_{2K_2}I^{2L_1-1}_{2K_3}}{\mu H+\varepsilon_4(L,L_1,K,K_3)}-\frac{I^{2K_3-1}_{2K_2}I^{2L_1-1}_{2K_3-1}}{\varepsilon_4(L,L_1,K,K_3)}\right)-\frac{I^{2K_2-1}_{2K_1}}{\varepsilon_4(L,L_1,K,K_2)}$$

$$\times\left(I^{2L_1-1}_{2K_2-1}-\frac{I^{2K_3}_{2K_2-1}I^{2L_1-1}_{2K_3}}{\mu H+\varepsilon_4(L,L_1,K,K_3)}+\frac{I^{2L_1-1}_{2L_2-1}I^{2L_2-1}_{2K_2-1}}{\varepsilon_4(L,L_2,K,K_2)}-\frac{I^{2K_3-1}_{2K_2-1}I^{2L_1-1}_{2K_3-1}}{\varepsilon_4(L,L_1,K,K_3)}\right)+\frac{I^{2L_1-1}_{2L_2}I^{2L_2}_{2L_3-1}I^{2L_3-1}_{2K_1}}{\varepsilon_4(L,L_2,K,K_1)(\mu H+\varepsilon_4(L,L_3,K,K_1))}$$

$$\left.-\frac{I^{2K_2}_{2L_2}I^{2L_2}_{2K_1}I^{2L_1-1}_{2K_2}}{(\mu H+\varepsilon_6)(\mu H+\varepsilon_4(L,L_1,K,K_2))}\right\}+\frac{I^{2K_1-1}_{2L_1-1}}{\varepsilon_4(L,L_1,K,K_1)-|I^{2K_2}_{2L_2-1}|^2/(\mu H+\varepsilon_6)-|I^{2K_2-1}_{2L_2-1}|^2/\varepsilon_6-\delta E}$$

$$\times\left\{I^{2L_1-1}_{2K_1-1}-\frac{I^{2K_2}_{2K_1-1}}{\mu H+\varepsilon_4(L,L_1,K,K_2)}\left(I^{2L_1-1}_{2K_2}-\frac{I^{2K_3}_{2K_2}I^{2L_1-1}_{2K_3}}{\mu H+\varepsilon_4(L,L_1,K,K_3)}-\frac{I^{2K_3-1}_{2K_2}I^{L_1-1}_{2K_3-1}}{\varepsilon_4(L,L_1,K,K_3)}\right)\right.$$

$$-\frac{I^{2K_2-1}_{2K_1-1}}{\varepsilon_4(L,L_1,K,K_2)}\left(I^{2L_1-1}_{2K_2-1}-\frac{I^{2K_3}_{2K_2-1}I^{2L_1-1}_{2K_3}}{\mu H+\varepsilon_4(L,L_1,K,K_3)}+\frac{I^{2L_1-1}_{2L_2-1}I^{2L_2-1}_{2K_2-1}}{\varepsilon_4(L,L_2,K,K_2)}-\frac{I^{2K_3-1}_{2K_2-1}I^{2L_1-1}_{2K_3-1}}{\varepsilon_4(L,L_1,K,K_3)}\right)$$

$$+\frac{I^{2L_1-1}_{2L_2-1}}{\varepsilon_4(L,L_2,K,K_1)}\left(I^{2L_2-1}_{2K_1-1}-\frac{I^{2K_2}_{2K_1-1}I^{2L_2-1}_{2K_2}}{\mu H+\varepsilon_4(L,L_2,K,K_2)}+\frac{I^{2L_2-1}_{2L_3-1}I^{2L_3-1}_{2K_1-1}}{\varepsilon_4(L,L_3,K,K_1)}-\frac{I^{2K_2-1}_{2K_1-1}I^{2L_2-1}_{2K_2-1}}{\varepsilon_4(L,L_2,K,K_2)}\right)$$

$$\left.+\frac{I^{2K_2-1}_{2L_2-1}}{\varepsilon_6}\left(\frac{I^{2L_1-1}_{2K_1-1}I^{2L_2-1}_{2K_2-1}}{\varepsilon_4(L,L_2,K,K_2)}-\frac{I^{2L_1-1}_{2K_2-1}I^{2L_2-1}_{2K_1-1}}{\varepsilon_4(L,L_2,K,K_1)}-\frac{I^{2L_2-1}_{2K_1-1}I^{2L_1-1}_{2K_2-1}}{\varepsilon_4(L,L_1,K,K_2)}\right)-\frac{I^{2K_2}_{2L_2-1}I^{2L_1-1}_{2K_2}I^{2L_2-1}_{2K_1-1}}{(\mu H+\varepsilon_6)\varepsilon_4(L,L_2,K,K_1)}\right\}.\tag{C2}$$

From Eqs. (6) and (A2), the quantities C_{2K}^{2L-1} and C_{2K-1}^{2L-1} can easily be obtained in the third order of perturbation theory. We do not give these expressions here because only one statement is essential for us: direct comparison of the quantities δE [Eq. (4)] and self-energy $\Sigma_{K,L}$ [Eq. (C2)] shows that

$$\delta E+\Sigma_{K,L}\big|_{\varepsilon_K=\varepsilon_L=\varepsilon_F}=0. \tag{C3}$$

Equation (C3) is valid for arbitrary spectrum $\varepsilon_K,\varepsilon_L$ and arbitrary transition matrix elements $I_{\cdots}^{\cdots}$. Our conjecture is that Eq. (C3) holds in all orders of perturbation theory, and hence we can put

$$\delta E+\Sigma_{K,L}\big|_{\varepsilon_k=\varepsilon_L=\varepsilon_F}=-\Delta, \tag{C4}$$

where Δ is exponentially small and can be considered as an order parameter. We also obtain from Eqs. (C1) and (C2) that self-energies $\Sigma_{KL}^{(1)}$ and Σ_{KL} coincide only in the second order of perturbation theory. They start to be different in the third order of perturbation theory. In the fourth order of perturbation theory, we obtain from Eqs. (C1) and (C2)

$$\begin{aligned}
\Sigma_{(K,L)}^{(1)}-\Sigma_{(K,L)}=&I_{2L_1}^{2K_1}I_{2L_2}^{2L_1}I_{2K_1}^{2L_2}\left(\frac{1}{(\mu H+\varepsilon_4(L,L_1,K,K_1)(\mu H+\varepsilon_4(L,L_2,K,K_1))}-\frac{1}{\varepsilon_4(L,L_1,K,K_1)\varepsilon_4(L,L_2,K,K_1)}\right)\\
&-I_{2K_1}^{2K_2}I_{2L_1}^{2K_1}I_{2L_2}^{2L_1}I_{2K_2}^{2L_2}\left\{\left(\frac{1}{\mu H+\varepsilon_4(L,L_1,K,K_1)}+\frac{1}{\varepsilon_4(L,L_1,K,K_1)}\right)\right.\\
&\times\left(\frac{1}{(\mu H+\varepsilon_4(L,L_2K,K_2))(\mu H+\varepsilon_4(L,L_1,K,K_2))}-\frac{1}{\varepsilon_4(L,L_1,K,K_2)\varepsilon_4(L,L_2,K,K_2)}\right)\\
&+\left(\frac{1}{\mu H+\varepsilon_4(L,L_2,K,K_2)}+\frac{1}{\varepsilon_4(L,L_2,K,K_2)}\right)\left(\frac{1}{(\mu H+\varepsilon_4(L,L_2,K,K_1))(\mu H+\varepsilon_4(L,L_1,K,K_1))}\right.\\
&\left.-\frac{1}{\varepsilon_4(L,L_2,K,K_1)\varepsilon_4(L,L_1,K,K_1)}\right)-\left(\frac{1}{\mu H+\varepsilon_4(L,L_2,K,K_1)}-\frac{1}{\varepsilon_4(L,L_2,K,K_1)}\right)\\
&\times\left(\frac{1}{(\mu H+\varepsilon_4(L,L_3,K,K_1))(\mu H+\varepsilon_4(L,L_1,K,K_1)}+\frac{1}{\varepsilon_4(L,L_3,K,K_1)\varepsilon_4(L,L_1,K,K_1)}\right)\\
&+\left(\frac{1}{\varepsilon_6(\mu H+\varepsilon_4(L,L_2,K,K_1))(\mu H+\varepsilon_4(L,L_1,K,K_1))}-\frac{1}{(\mu H+\varepsilon_6)\varepsilon_4(L,L_2,K,K_1)\varepsilon_4(L,L_1,K,K_1)}\right)\\
&+\left(\frac{1}{(\mu H+\varepsilon_6)(\mu H+\varepsilon_4(L,L_1,K,K_1)}\right)\left(\frac{1}{\mu H+\varepsilon_4(L,L_2,K,K_1)}-\frac{1}{\mu H+\varepsilon_4(L,L_2,K,K_2)}\right)\\
&\left.-\left(\frac{1}{\varepsilon_4(L,L_2,K,K_1)}-\frac{1}{\varepsilon_4(L,L_2,K,K_2)}\right)\frac{1}{\varepsilon_6\varepsilon_4(L,L_1,K,K_1)}\right\},
\end{aligned} \tag{C5}$$

where

$$\varepsilon_4(L,L_1,K,K_1)\equiv\varepsilon_L+\varepsilon_{L_1}-\varepsilon_K-\varepsilon_{K_1},$$

$$\varepsilon_6\equiv\varepsilon_L+\varepsilon_{L_1}+\varepsilon_{L_2}-\varepsilon_K-\varepsilon_{K_1}-\varepsilon_{K_2}. \tag{C6}$$

Straightforward calculation of the integrals in Eq. (C5) leads to Eqs. (40) and (41). Both Eqs. (40) and (41) are proved in two orders of perturbation theory. Our conjecture is that Eq. (41) is exact.

APPENDIX D

In this Appendix we consider the role of the right-hand side of Eqs. (7) for a negative value of the coupling constant, $g<0$. In the first order of perturbation theory, we obtain from (A2)

$$\begin{aligned}
C_{2K_1;2K}^{2L_1;2L-1}=&\frac{1}{\mu\tilde{H}+\varepsilon_4(L,L_1,K,K_1)+\Delta}\\
&\times[C_{2K_1}^{2L-1}I_{2K}^{2L_1}-C_{2K}^{2L-1}I_{2K_1}^{2L_1}+C_{2K}^{2L_1}I_{2K_1}^{2L-1}\\
&-C_{2K_1}^{2L_1}I_{2K}^{2L-1}],
\end{aligned}$$

$$C^{2L_1;2L-1}_{2K_1-1;2K}=\frac{1}{\varepsilon_4(L,L_1,K,K_1)+\Delta}\times[I^{2L_1}_{2K_1-1}C^{2L-1}_{2K}-C^{2L_1}_{2K}I^{2L-1}_{2K_1-1}],$$

$$C^{2L_1-1;2L-1}_{2K_1;2K-1}=\frac{1}{\mu\tilde{H}+\varepsilon_4(L,L_1,K,K_1)+\Delta}\times[C^{2L_1-1}_{2K-1}I^{2L-1}_{2K_1}-C^{2L-1}_{2K-1}I^{2L_1-1}_{2K_1}],$$

$$C^{2L_1-1;2L-1}_{2K_1-1;2K-1}=\frac{1}{\varepsilon_4(L,L_1,K,K_1)+\Delta}[I^{2L_1-1}_{2K_1-1}C^{2L-1}_{2K-1}-C^{2L-1}_{2K_1-1}I^{2L_1-1}_{2K-1}+C^{2L_1-1}_{2K_1-1}I^{2L-1}_{2K-1}-C^{2L_1-1}_{2K-1}I^{2L-1}_{2K_1-1}]. \quad \text{(D1)}$$

Inserting (D1) into (6), we obtain

$$A_1(C^{2L-1}_{2K};C^{2L-1}_{2K-1},C^{2L}_{2K})=-\sum\frac{I^{2K_1}_{2L_1}(C^{2L-1}_{2K_1}I^{2L_1}_{2K}+C^{2L_1}_{2K}I^{2L-1}_{2K_1}-C^{2L_1}_{2K_1}I^{2L-1}_{2K})}{\mu\tilde{H}+\varepsilon_4(L,L_1,K,K_1)+\Delta}-\sum\frac{I^{2K_1-1}_{2L_1}I^{2L-1}_{2K_1-1}C^{2L_1}_{2K}}{\varepsilon_4(L,L_1,K,K_1)+\Delta},$$

$$A_2(C^{2L-1}_{2K};C^{2L-1}_{2K-1};C^{2L}_{2K})=-\sum\frac{I^{2K_1-1}_{2L_1-1}(C^{2L-1}_{2K_1-1}I^{2L_1-1}_{2K-1}-C^{2L_1-1}_{2K_1-1}I^{2L-1}_{2K-1}+C^{2L_1-1}_{2K-1}I^{2L-1}_{2K_1-1})}{\varepsilon_4(L,L_1,K,K_1)+\Delta}-\sum\frac{I^{2K_1}_{2L_1-1}I^{2L-1}_{2K_1}C^{2L_1-1}_{2K-1}}{\mu\tilde{H}+\varepsilon_4(L,L_1,K,K_1)+\Delta},$$

$$A_3(C^{2L-1}_{2K};C^{2L-1}_{2K-1},C^{2L}_{2K})=\sum\frac{I^{2K_1}_{2L_1-1}(C^{2L_1-1}_{2K_1}I^{2L}_{2K}-C^{2L_1-1}_{2K}I^{2L}_{2K_1}-C^{2L}_{2K_1}I^{2L_1-1}_{2K})}{\mu\tilde{H}+\varepsilon_4(L,L_1,K,K_1)+\Delta}-\sum\frac{I^{2K_1-1}_{2L_1-1}I^{2L}_{2K_1-1}C^{2L_1-1}_{2K}}{\varepsilon_4(L,L_1,K,K_1)+\Delta}. \quad \text{(D2)}$$

The quantities $\varepsilon_4,\varepsilon_6$ here are the same as in Eq. (C6).

As before, only convolutions Z_L,Y_L are large for $g<0$. Furthermore,

$$|Z_L+Y_L|\sim g^2|Z_L-Y_L|. \quad \text{(D3)}$$

As the result, Eqs. (7) can be reduced to just one equation:

$$(Z_L-Y_L)\left[1+g\ln\frac{\varepsilon_F}{y+\Delta}+g\ln\frac{\varepsilon_F}{\mu\tilde{H}+y+\Delta}+\frac{g^3}{2}\left(\frac{I_1}{g\ln[\varepsilon_F/(\mu\tilde{H}+y+\Delta)]}+\frac{I_2}{g\ln[\varepsilon_F/(y+\Delta)]}\right)\right]$$

$$=Ig\left(\ln\frac{\varepsilon_f}{\mu\tilde{H}+y+\Delta}+\ln\frac{\varepsilon_F}{y+\Delta}\right)+g\int dx\left(\frac{X_K}{\mu\tilde{H}+y+x+\Delta}-\frac{Y_K}{y+x+\Delta}\right), \quad \text{(D4)}$$

where

$$I_1=\int\frac{dx\,dy\,dx_1}{(\mu\tilde{H}+y+x+\Delta)(\mu\tilde{H}+y+x_1+\Delta)(\mu\tilde{H}+y+x+y_1+x_1+\Delta)},$$

$$I_2=\int\frac{dx\,dy\,dx_1}{(y+x+\Delta)(y+x_1+\Delta)(y+x+y_1+x_1+\Delta)}. \quad \text{(D5)}$$

A simple calculation of the integrals (D5) gives

$$I_1=\frac{1}{3}\ln^3\left(\frac{\varepsilon_F}{\mu\tilde{H}+y+\Delta}\right),$$

$$I_2=\frac{1}{3}\ln^3\left(\frac{\varepsilon_F}{y+\Delta}\right). \tag{D6}$$

Now we can define the Kondo temperature T_c to be

$$|g|\ln\frac{\varepsilon_F}{T_c}=z, \tag{D7}$$

where z is a root of the quadratic equation

$$1-2z+\frac{z^2}{3}=0, \quad z=3-\sqrt{6}\approx 0.5505. \tag{D8}$$

From Eq. (D4) we obtain

$$Z_L-Y_L=\frac{I\tilde{\beta}}{|g|(1-z/3)\ln((\mu\tilde{H}+y+\Delta)(y+\Delta)/T_c^2)}, \tag{D9}$$

where $\tilde{\beta}$ is a number of order 1. Instead of Eqs. (41) and (42) we have now

$$\mu\tilde{H}=\mu H-\delta\Sigma; \quad \delta\Sigma=-\mu Hz(-1/2+\langle S_z\rangle). \tag{D10}$$

As before, the average spin $\langle S_z\rangle$ is given by Eq. (27) with the replacement $\mu H\to\mu\tilde{H}$:

$$\langle S_z\rangle=\frac{\mu\tilde{H}}{4(T_c^2+(\mu\tilde{H}/2)^2)^{1/2}}. \tag{D11}$$

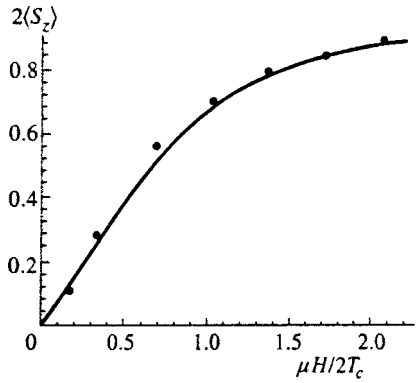

FIG. 2. Magnetic field dependence of the average spin $\langle S_z\rangle$. Dots are experimental of Ref. 4.

The magnetic field dependence of the average spin $\langle S_z\rangle$ [Eqs. (D10) and (D11)] is given in Fig. 2. Dots are the experimental results of Ref. 4.

[1] A. A. Abrikosov and A. A. Migdal, J. Low Temp. Phys. **3**, 519 (1970).
[2] A. M. Tsvelick and P. B. Wigmann, Adv. Phys. **32**, 453 (1983).
[3] N. Andrei, K. Furuya, and J. H. Lowenstein, Rev. Mod. Phys. **55**, 331 (1983).
[4] W. Felsch, Z. Phys. B **29**, 212 (1978).
[5] S. D. Bader, N. E. Phillips, M. B. Haple, and C. A. Luengo, Solid State Commun. **16**, 1263 (1975).
[6] Yu. N. Ovchinnikov, A. M. Dyugaev, P. Fulde, and V. Z. Kresin, JETP Lett. **66**, 184 (1997).
[7] P. Nozieres and T. C. de Dominicis, Phys. Rev. **178**, 1097 (1969).
[8] K. Yosida, Phys. Rev. **147**, 223 (1966).
[9] Hiroumi Ishii, Prog. Theor. Phys. **40**, 201 (1968).
[10] Hiroumi Ishii, Prog. Theor. Phys. **43**, 578 (1970).
[11] M. Fowler and A. Zawadowskii, Solid State Commun. **9**, 471 (1971).

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

Random walk on hierarchical comb structures

V. E. Arkhincheev*)

Buryat Science Center, Siberian Branch of the Russian Academy of Sciences, 670047 Ulan-Ude, Russia
(Submitted 2 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 1285–1296 (April 1999)

This paper examines random walks on an exactly solvable comb model of percolation clusters. The study shows that diffusion along the structure's axis is anomalous. Generalized diffusion equations with fractional-order time derivatives are derived, and a generalization to the multidimensional case is carried out. The relationship between this problem and that of diffusion in a medium with traps is examined, and equations that describe diffusion in a medium with traps are derived. The paper also discusses the transition to ordinary diffusion due to the introduction of comb teeth of finite length, and analyzes the case of N teeth of different length. It is shown that the solution of this problem leads to the emergence of an N-channel diffusion equation. Finally, equations describing the diffusion of interacting electrons are derived.
 [S1063-7761(99)00904-X]

1. INTRODUCTION

Interest in random walks in highly inhomogeneous media and along fractals can be explained both by the numerous applications of diffusion problems, e.g., the problem of conductivity in highly inhomogeneous media, and by the unusual anomalous nature of random walks along fractals. The anomalous nature manifests itself in the power-function dependence of the mean-square displacement on time:[1]

$$\langle X^2(t)\rangle \propto t^{2/(2+\theta)}, \quad \theta \geqslant 0. \tag{1}$$

The law (1) was established by the renormalization-group method for regular fractals of the Sierpinski tiling type[1,2] and by computer simulation for percolation clusters, or statistical fractals.[3,4] There are at least two reasons for the change in the nature of diffusion: the marked sinuosity of percolation paths on all scales, and the existence of dead ends in the current-carrying paths. Weiss and Havlin[5] proposed a model that allows for the existence of dead ends in percolation systems, a comb structure (Fig. 1). They used the generating-function technique to demonstrate that the time dependence of the mean-square displacement along the structure's axis is of the anomalous type (1) and that $\theta=2$. However, they did not derive an equation and proposed an incorrect extrapolation expression of the Gaussian type for the Green's function. In Refs. 6 and 7 a rigorous description of diffusion on such a structure was given and a diffusion equation describing random walks along the structure's axis was derived. This equation differs from a continuity equation in that instead of a first-order time derivative it contains a fractional derivative of order 1/2. (The expression for the diffusion current J is the ordinary one.) Attention in these papers was focused on demonstrating the effect of an electric field on diffusion and on establishing the relationship between diffusion and conductivity for the anomalous case.

The comb structure model is one of the few exactly solvable models with unusual diffusion properties. The model explicitly allows for the effect of dead ends on the nature of diffusion. Therefore, further investigations into random walks along a comb structure are of interest. (The author hopes that the results of such an investigation will reflect the special features of diffusion processes in real media of the percolation type.) The present paper studies the various hierarchical structures in the given model and generalizes the results to the multidimensional case. It also studies the transition to ordinary diffusion when the teeth of the comb structure are of finite length. An equation is derived that describes a random walk on a comb structure with teeth of finite length. It appears that the most interesting case involves diffusion on a comb structure in which the N teeth are of various lengths. Here the random walk largely depends on which teeth the particle visited and on the length of each of these teeth. The solution of this problems leads an N-channeldiffusion equation, which connects all N teeth. Asymptotic solutions of this equation are obtained. Another problem studied in the paper is that of diffusion in media with traps (continuous-time random walk). A system of equations for this problem is derived. Finally, equations that describe the diffusion of interacting electrons are set up, and solutions of these equations are found. The results are discussed in the Conclusion.

2. DIFFUSION ON A COMB STRUCTURE

Let us briefly recall the results of Ref. 6. A specific feature of diffusion in the adopted model is that a displacement in the x direction is only possible along the structure's axis at $y=0$. In other words, the diffusion coefficient D_{xx} is nonzero only at $y=0$:

$$D_{xx}=D_1\delta(y)\ (J_x=-D_1\delta(y)\,\partial^2\rho/\partial x^2), \tag{2}$$

where ρ is the number density of the diffusing particles. Diffusion along the teeth is assumed to be ordinary: $D_{yy}=D_2$. Thus, a random walk on a comb structure is described by the diffusion tensor

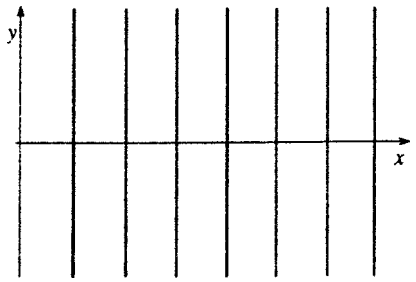

FIG. 1. Comb structure: infinitely long ''teeth'' are attached to a conducting axis ($y=0$).

$$\hat{D}=\begin{pmatrix} D_1\delta(y) & 0 \\ 0 & D_2 \end{pmatrix}. \tag{3}$$

Accordingly, we obtain the following diffusion equation:

$$\left(\frac{\partial}{\partial t}-D_1\delta(y)\frac{\partial^2}{\partial x^2}-D_2\frac{\partial^2}{\partial y^2}\right)G(x,y,t)=\delta(x)\,\delta(y)\,\delta(t), \tag{4}$$

where $G(x,y,t)$ is the Green's function of the diffusion problem. Using the Laplace transform with respect to time and the Fourier transform with respect to position, we obtain a mixed (s,k,y)-representation of this equation:

$$(s+D_1k^2\delta(y)-D_2\,\partial^2/\partial y^2)G(s,k,y)=\delta(y). \tag{5}$$

The solution of Eq. (5) is

$$G(s,k,y)=\frac{\exp(-\sqrt{s/D_2}\,|y|)}{2\sqrt{sD_2}+D_1k^2}. \tag{6}$$

We study diffusion along the structure's axis, i.e., at $y=0$. The corresponding Green's function is

$$G(x,0,t)=\frac{1}{2\pi\sqrt{D_1t^3}}\int_0^\infty \exp\left(-\frac{x^2}{4D_1\tau}-\frac{D_2\tau^2}{4t}\right)\sqrt{D_2\tau}\,d\tau. \tag{7}$$

In deriving (7) we used the identity $\int\exp(-\alpha\tau)\,d\tau=1/\alpha$. The total number of particles on the structure's axis decreases with time: $\langle G\rangle=\int G(x,0,t)\,dx=1/(2D_2t)$. Thus, the Green's function $G(x,0,t)$ describes diffusion with a nonconserved number of particles, because the particles leave for teeth of infinite length. Allowing for this fact, we calculate the displacement along the structure's axis:

$$\langle X^2(t)\rangle=\frac{\langle X^2G\rangle}{\langle G\rangle}=D_1\sqrt{\frac{t}{D_2}}. \tag{8}$$

Let us discuss the equation for $G(x,0,t)$. From (6) it follows that in the (s,k)-representation the equation is

$$[2(sD_2)^{1/2}+D_1k^2]\rho(s,k)=0. \tag{9}$$

Using the definition of fractional-order time derivative,[8] we obtain a diffusion equation for the particle number density on the structure's axis:

$$\left[\frac{\partial^{1/2}}{\partial t^{1/2}}+\frac{D_1}{D_2^{1/2}}\frac{\partial^2}{\partial x^2}\right]\rho(x,t)=0, \tag{10}$$

where the operator of fractional differentiation with respect to time is

$$\frac{\partial^{1/2}f}{\partial t^{1/2}}=\int_{-\infty}^{\infty}\frac{\partial f(\tau)}{\partial\tau}\frac{d\tau}{|t-\tau|^{1/2}}.$$

The fact that the diffusion equation (10) is of an integro-differential type is a corollary of random disappearance and subsequent emergence of particles (the departure of particles from the axis and their return) as they wander along the structure's axis. Our goal here is a further generalization of these results.

3. MULTIDIMENSIONAL CASE

Let us first examine a three-dimensional comb structure. Such a structure is formed by attaching additional teeth to the existing comb structure that point in the direction parallel to the z axis. Hence in the three-dimensional case displacements in the x-direction are possible only along the intersection of the planes $y=0$ and $z=0$. In other words, the diffusion coefficient D_{xx} is finite only if $y=0$ and $z=0$, i.e., $D_{xx}=D_1\delta(y)\,\delta(z)$. Accordingly, a displacement in the y-direction is possible only if $z=0$, and displacement along the z axis is ordinary. Thus, we obtain the diffusion tensor

$$\hat{D}=\begin{pmatrix} D_1\delta(y)\,\delta(z) & 0 & 0 \\ 0 & D_2\delta(z) & 0 \\ 0 & 0 & D_3 \end{pmatrix}, \tag{11}$$

so that the corresponding diffusion equation with coefficient (11) in the mixed (s,k,y,z)-representation is

$$\left[s-D_1k^2\delta(y)\,\delta(z)-D_2\delta(z)\frac{\partial^2}{\partial y^2}-D_3\frac{\partial^2}{\partial z^2}\right]\rho(s,k,y,z)=0. \tag{12}$$

We seek a solution of Eq. (12) in the form

$$\rho(s,k,y,z)=g(s,k)\,\exp\,(-\lambda_2|y|-\lambda_3|z|). \tag{13}$$

Substituting (13) into Eq. (12) yields the following formulas for the parameters λ_2 and λ_3 and the function $g(s,k)$:

$$\lambda_3^2=\frac{s}{D_3},\quad \lambda_2^2=\frac{2\lambda_3D_3}{D_2}=\frac{2(sD_3)^{1/2}}{D_2}, \tag{14}$$

$$g(s,k)=\frac{1}{2\lambda_2D_2+D_1k^2}. \tag{15}$$

For the mean-square displacement along the x and y axes we then have

$$\langle X^2(t)\rangle\propto t^{1/4}, \tag{16}$$

$$\langle Y^2(t)\rangle\propto t^{1/2}. \tag{17}$$

Hence in the N-dimensional case the diffusion tensor is described by the matrix

$$\hat{D}=\begin{pmatrix} D_1\delta(x_2)\cdots\delta(x_N) & 0 & \cdots & 0 & 0 \\ 0 & D_2\delta(x_3)\cdots\delta(x_N) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & D_{N-1}\delta(x_N) & 0 \\ 0 & 0 & \cdots & 0 & D_N \end{pmatrix}. \tag{18}$$

Accordingly, we seek a solution of the N-dimensional diffusion problem in the form

$$\rho(s,k,x_2,x_3,\ldots,x_N) = g(s,k)\ \exp(-\lambda_2|x_2|-\lambda_3|x_3|-\cdots-\lambda_N|x_N|). \tag{19}$$

Here the parameters λ_n are linked through the formulas

$$2\lambda_N=\frac{s}{D_N},\quad \lambda_{N-1}^2=\frac{2\lambda_N D_N}{D_{N-1}},\ldots,\quad \lambda_2^2=\frac{2\lambda_3 D_3}{D_2}, \tag{20}$$

and the function $g(s,k)$ is defined in (15). The expressions (19) and (20) comprise the complete solution of the problem. For instance, we can easily calculate the mean-square displacement along the main axis of the structure:

$$\langle X_N^2(t)\rangle\propto t^{1/2(N-1)}. \tag{21}$$

Now, the mean-square displacement for the next lateral tooth (this tooth and the attached teeth comprise an $(N-1)$-dimensional structure) is

$$\langle X_{N-1}^2(t)\rangle\propto t^{1/2(N-2)}, \tag{22}$$

etc. On the penultimate axis, from which only teeth of infinite length emerge, we have

$$\langle X_2^2(t)\rangle\propto t^{1/2}. \tag{23}$$

Thus, a random walk on a multidimensional comb structure is of a hierarchical nature, and there are many variants of behavior of the mean-square displacement along the axes of the structure. Here the equations describing a random walk along the mth axis of an N-dimensionalstructure can be represented in the form of a system of equations in which the first-order time derivative is replaced with a fractional derivative of the appropriate order:

$$(s^\alpha+\partial^2/\partial x_m^2)\rho(s,x)=0, \tag{24}$$

where $\alpha=1/2(N-m)$.

4. COMB STRUCTURE WITH TEETH OF FINITE LENGTH

Up to this point we examined comb structures with infinitely long teeth. Now we turn to the case of diffusion along a two-dimensional structure with teeth of finite length L and reflecting boundaries. The following method will be used to solve this problem.

We write Eq. (4) in the form of an ordinary diffusion equation with an inhomogeneous right-hand side,

$$\frac{\partial\rho}{\partial t}-D_2\frac{\partial^2\rho}{\partial y^2}=D_1\delta(y)\frac{\partial^2\rho}{\partial x^2}, \tag{25}$$

and the boundary conditions

$$J(y=\pm L)=0. \tag{26}$$

The Green's function of Eq. (25) with the boundary conditions (26) is well known:

$$G(y,t)=\frac{1}{L}\sum_{m=0}^{\infty}\ \exp\left(-D_2 t\frac{m^2\pi^2}{L^2}\right)\cos\frac{m\pi y}{L}. \tag{27}$$

Thus we obtain an integral equation for the concentration:

$$\rho(x,y,t)=\int G(y-y',t-t')D_1\delta(y')\frac{\partial^2\rho(x,y',t')}{\partial x^2}dy'dt'.$$

This equation has the simplest form in the (s,k,y)-representation:

$$\rho(s,k,y)=-\frac{D_1k^2}{L}\sum\ \frac{\cos(m\pi y/L)}{s+m^2\pi^2/L^2}\rho(s,k,0). \tag{28}$$

At $y=0$ we obtain a closed-form equation for $\rho(s,k,0)$:

$$K(s,L)\rho(s,k)=-D_1k^2\rho(s,k). \tag{29}$$

Here the inverse operator is

$$K^{-1}=\frac{1}{sL}+\frac{\coth[L(sD_2)^{1/2}]}{2(sD_2)^{1/2}}. \tag{30}$$

When the teeth are of infinite length ($L=\infty$), we have the well-known formula $K(s,\infty)=2(sD_2)^{1/2}$. For long time intervals we have the ordinary asymptotic equation for diffusion with the diffusion coefficient depending on tooth length:

$$[s+\text{const}\times D_1k^2/L]\rho(s,k)=0. \tag{31}$$

The structure that we have studied had teeth of equal length L. Now we assume that the N teeth have various lengths, $L_1,L_2,\ldots,L_N$, and that this pattern repeats periodically. The distance between the sites on the structure's axis is a. To understand how a random walk on such a structure can be described, we analyze the case of two lengths, L_1 and L_2. We write the second derivative with respect to the coordinate x in the finite-difference form and introduce the notation $K(s,L=L_1)=K_1$ and $K(s,L=L_2)=K_2$. We also denote the particle concentration at the point on the axis to which a tooth of length L_1 is attached by F_1 (F_2 is intro-

duced in a similar way). Then the following system of equations describing the behavior of the particles on the axis can be written:

$$K_1F_1(x)=D_1\frac{F_2(x+a)+F_2(x-a)-2F_1(x)}{a^2},$$

$$K_2F_2(x)=D_1\frac{F_1(x+a)+F_1(x-a)-2F_2(x)}{a^2} \tag{32}$$

or, in the k-representation,

$$\begin{pmatrix} K_1(s)-2\dfrac{D_1}{a^2} & 2D_1\cos ka \\ 2D_1\cos ka & K_2(s)-2\dfrac{D_1}{a^2}\end{pmatrix}\begin{pmatrix}F_1(k)\\F_2(k)\end{pmatrix}=0. \tag{33}$$

Setting the determinant of this equation to zero, we can find the relationship between the parameters s and k, or in other words, the analog of the diffusion equation in the (s,k)-representation:

$$T_1(s)T_2(s)-C^2\cos^2 ka=0, \tag{34}$$

where $T(s)=K(s)-C$, and $C=2D_1/a^2$. From Eq. (34) with equal tooth lengths and as $a\to 0$ we obtain Eq. (9), as expected.

Thus, to describe random walks on N teeth of differing length, we must set up a system of N equations. Such a system emerges because diffusion strongly depends on what teeth (and of what length) are involved in the random walk of a particle. The above analysis suggests that in the case of a comb structure with N teeth, the determinant takes the form

$$\begin{pmatrix}
T_1 & \exp(ika) & 0 & 0 & \cdots & 0 & \exp(-ika)\\
\exp(-ika) & T_2 & \exp(ika) & 0 & \cdots & 0 & 0\\
0 & \exp(-ika) & T_3 & \exp(ika) & \cdots & 0 & 0\\
0 & 0 & \exp(-ika) & T_4 & \cdots & \vdots & \vdots\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
\exp(ika) & 0 & 0 & \cdots & \cdots & \exp(-ika) & T_N
\end{pmatrix}=0. \tag{35}$$

Thus, instead of the ordinary diffusion equation we have an N-channeldiffusion equation, i.e., instead of a fairly simple dispersion law $s=k^2$, valid for ordinary diffusion, we have an Nth order equation with N solutions. Moreover, according to (30), the type of operator K depends on the relationship between the parameter s and the diffusion time $t_i=D/L_i^2$ along a tooth.

Let us analyze the solutions of Eqs. (35) by qualitative reasoning. Suppose the tooth lengths differ substantially. Then the problem acquires a hierarchy of times related to diffusion along these teeth: $t_1\ll t_2\ll t_3\ll\cdots\ll t_N$. Over short times, shorter than any characteristic times of the problem, diffusion is anomalous, as it is in a comb structure with infinitely long teeth. As time increases, anomalous diffusion is replaced by ordinary diffusion with a diffusion coefficient depending on the length of the particular tooth. Over times $t_1\ll t\ll t_2$, $D\sim D_1/L_1$; over times $t_2\ll t\ll t_3$, $D\sim D_1/L_2$; etc.:

$$\langle X^2(t)\rangle\propto t^{1/2},\quad t\ll t_1\ll\cdots\ll t_N,$$

$$\langle X^2(t)\rangle\propto D_1t/L_m,\quad t_m\ll t\ll t_{m+1}.$$

A more detailed solution of this system merits a separate investigation.

5. CONTINUOUS-TIME RANDOM WALK

The above problem of a random walk on an N-dimensional comb structure is related to the problem of diffusion in a medium with traps (continuous-time random walk). What makes the two problems different is that in diffusion in a medium with traps the particles do not disappear but, with a certain probability, stay at each site. The total number of diffusing particles is conserved.[9,10] For a comb structure, the statement of the problem with a continuous distribution over the time lags in the two-dimensional case amounts to studying the quantity

$$\tilde G(x,t)=\int G(x,y,t)\,dy. \tag{36}$$

According to (5), the function $\tilde G(x,t)$ is described by the equation

$$\left[s+\frac{D_1k^2s^{1/2}}{D_2}\right]\tilde G=1. \tag{37}$$

Hence, in the case of a medium with traps, the diffusion equation has the form of the continuity equation for a medium with temporal dispersion:

$$\frac{\partial\rho(x,t)}{\partial t}-\frac{\partial J}{\partial x}=0, \tag{38}$$

where

$$J=-\frac{D_1}{2D_2}\frac{\partial}{\partial x}\int\frac{\partial\rho(x,\tau)}{\partial\tau}\frac{d\tau}{|t-\tau|^{1/2}}. \tag{39}$$

Diffusion is still anomalous with the same exponent $\theta=2$.

We now turn to the three-dimensional case and examine $G(s,k,y,a)$ averaged over the y and z axes, i.e., the function $\tilde{G}(s,k)=\int G(s,k,y,z)\,dy\,dz$. According to (13), for this function we have the equation

$$\left[sD_2+D_1k^2\left(\frac{4sD_3}{D_2}\right)^{3/4}\right]\tilde{G}=1. \tag{40}$$

Hence the diffusion equation has the form of the continuity equation with a diffusion current

$$J(x,t)\propto-\frac{\partial}{\partial x}\int\frac{\partial\rho(x,\tau)}{\partial\tau}\frac{\partial\tau}{|t-\tau|^{3/4}}. \tag{41}$$

Now we study the Green's function averaged over one coordinate, z:

$$\tilde{G}(k,y,t)=\exp\,(-\lambda_2|y|)/\lambda_3(2\lambda_2D_2+D_1k^2)\,. \tag{42}$$

Accordingly, motion along the $y=0$ axis is described by the equation

$$[s^{3/4}+As^{1/2}k^2]\tilde{G}_1(s,k)=0,\quad A=\text{const}. \tag{43}$$

The number of particles on the $y=0$ axis is not conserved, since they leave for the dead ends along the $y=0$ axis. The diffusion current also contains a time derivative of order 1/2.

Clearly, in the N-dimensional case the equation for the function $\tilde{G}_m$ averaged over the m coordinates has the form

$$(s^{\beta}+s^{\nu}k^2)\tilde{G}_m(s,k)=0, \tag{44}$$

where $\beta=(N-m+1)/4$, and $\nu=(N-m-1)/4$.

6. DIFFUSION ON A COMB STRUCTURE OF INTERACTING ELECTRONS

Generally, the current has a diffusion component and a field component:

$$\mathbf{j}=-D\nabla\rho+\sigma\mathbf{E}, \tag{45}$$

where the diffusion coefficient D and the conductivity σ have the tensor form (3), and the electric field satisfies Poisson's equation

$$\text{div}\ \mathbf{E}=4\pi\rho. \tag{46}$$

Hence the distribution of electric potential over the comb structure is described by the equation for the 4-potential with singular coefficients:

$$\left\{\left[\frac{\partial}{\partial t}-D_1\delta(y)\frac{\partial^2}{\partial x^2}-D_2\frac{\partial^2}{\partial y^2}\right]\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)-\sigma_1\delta(y)\times\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\right\}\varphi=0. \tag{47}$$

Using the Laplace transform with respect to time and the Fourier transform with respect to x, we obtain an equation for the potential in the mixed (s,k,y)-representation:

$$\left\{\left[s+D_1k^2\delta(y)-D_2\frac{\partial^2}{\partial y^2}\right]\left(-k^2+\frac{\partial^2}{\partial y^2}\right)+4\pi\sigma_1\delta(y)k^2-4\pi\sigma_2\frac{\partial^2}{\partial y^2}\right\}\varphi=0. \tag{48}$$

To find the general form of the solution of Eq. (48), we first examine a pure diffusion problem ($\sigma_1=\sigma_2=0$) and then find the potential for this case. Such an approach also suggests a way of solving Eq. (48) in the general case. According to (6), the number density of diffusing particles on the structure is given by

$$\rho(s,k,y)=\rho(s,k)\exp(-\lambda|y|), \tag{49}$$

i.e., in the (s,k,y)-representation we have

$$\rho(s,k,y)=\rho(s,k)\int\exp\,(-\lambda|y|+iqy)\,dy=\rho(s,k)\frac{2\lambda}{\lambda^2+q^2}. \tag{50}$$

Thus, according to Poisson's equation, the electric potential φ is given by

$$\varphi(s,k,q)=\frac{A(s,k)}{(\lambda^2+q^2)(k^2+q^2)}. \tag{51}$$

If we again transform to the mixed (s,k,y)-representation, we find that

$$\varphi(s,k,y)=\phi(s,k)\left[\frac{\exp(-k|y|)}{k}+\frac{\exp(-\lambda|y|)}{\lambda}\right]. \tag{52}$$

Note that after differentiating this expression twice with respect to y, the singular parts cancel. Hence, we seek a solution of Eq. (48) in a similar form:

$$\varphi(s,k,y)=\phi(s,k)\left[\frac{\exp(-\mu|y|)}{\mu}+\frac{\exp(-\lambda|y|)}{\lambda}\right]. \tag{53}$$

The parameters μ and λ can be found by substituting the solution (53) into Eq. (48):

$$\mu^2=\frac{s+4\pi\sigma+D_2k^2-[(s+4\pi\sigma+D_2k^2)^2+4D_2sk^2]^{1/2}}{2D_2},$$
$$\lambda^2=\frac{s+4\pi\sigma+D_2k^2+[(s+4\pi\sigma+D_2k^2)^2+4D_2sk^2]^{1/2}}{2D_2}. \tag{54}$$

The signs in (54) are determined by the condition that at $\sigma_1=\sigma_2=0$ the solution (53) becomes (52), i.e.,

$$\lim_{\substack{\sigma_1\to0\\ \sigma_2\to0}}\mu^2=k^2,\quad \lim_{\substack{\sigma_1\to0\\ \sigma_2\to0}}\lambda^2=\frac{s}{D_2}. \tag{55}$$

Accordingly, we have an expression for the function $\phi(s,k)$:

$$\phi(s,k)=\left[\frac{(2\mu D_2+D_1k^2)(\mu^2-k^2)-\sigma_1k^2}{\mu}-\frac{(2\lambda D_2+D_1k^2)(\lambda^2-k^2)-\sigma_1k^2}{\lambda}\right]^{-1}. \quad (56)$$

Formulas (53)–(56) provide a complete solution of the problem of diffusion in interacting particles in a comb structure. Heretofore, the diffusion of interacting particles in inhomogeneous media has been studied only by computer simulation.

How does the general solution (56) of the problem become the solution of the relaxation problem at $D_1=D_2=0$? To answer this question, we note that in this passage to the limit the quantities that remain constant are

$$\lim_{D_2\to 0}\mu^2=k^2\frac{s}{s+4\pi\sigma_2},$$
$$\lim_{D_2\to 0}D_2(\lambda^2-k^2)=s+4\pi\sigma_2, \quad (57)$$

and the parameter λ tends to infinity as $1/D_2^{1/2}$. With allowance for this fact, we can easily use (53)–(56) to derive an expression for the electric potential in the problem of the spreading of electric charge in a comb structure:

$$\varphi(s,k,y)=\frac{\exp\{-k|y|[s/(s+4\pi\sigma_2)]\}}{2k[s(s+4\pi\sigma_2)]^{1/2}+4\pi\sigma_1k^2}. \quad (58)$$

7. CONCLUSION

We have studied random walks in the comb-structure model and found that the existence of dead ends in the current-carrying paths, teeth in the comb structure, leads to the anomalous nature of the random walk. We have established that for diffusion problems in which the number of particles is not conserved the generalized diffusion equation must be of the integro-differential type: instead of having a first time derivative, the equation must contain a fractional-order derivative [see Eq. (24)]. Fractional-order time derivatives emerge because of the random disappearance and reappearance of diffusing particles (the departure of particles from the axis and their return).

In the case of teeth of finite length, over long times anomalous diffusion is replaced by ordinary diffusion, but the diffusion coefficient is explicitly dependent on the tooth length L. It is shown that a random walk on multidimensional comb structures with a variety of tooth lengths is of a hierarchical nature and that different power functions representing the time dependence of the mean-square displacement are possible. Thus, to explicitly allow for the effect of dead ends in the percolation paths in inhomogeneous media, one must use generalized equations with fractional-order time derivatives.

The situation is completely different when we examine random walks in a medium with a continuous distribution of time lags on traps. As noted earlier, the problem of diffusion in a medium with traps differs from the problem of diffusion along the axis of a comb structure with departure to dead ends. The difference lies in the fact that the particles do not disappear but, with a certain probability, stay at each site. The total number of diffusing particles is conserved. Hence we have the law of mass conservation, expressed by a continuity equation. However, the anomalous nature of diffusion, due to the capture of particles by the traps, leads to an unusual expression for the diffusion current [see Eq. (39)]. Note that mathematically Eqs. (24) and (38) with current (39) are different. Despite the fact that in both problems diffusion is anomalous, has the same exponent $\theta=2$, and the solutions of the respective equations are almost the same, these equations describe different physical situations. First, in diffusion along the axis of a comb structure the number of particles is not conserved. Second, the diffusion fluxes are different.

The equations with fractional-order time derivative similar to those discussed in the present paper appear in the description of diffusion on Cantor sets.[11] In their review,[12] Olemskoĭ and Flat discuss these results and interpret fractional-order time derivatives. According to Ref. 12, the fractional order of the time derivative corresponds to the relative fraction of mechanical (reversible in time) and dissipative diffusion (irreversible) processes. They do not really distinguish between Eqs. (24) and (38) with allowance for (39), i.e., between fractional-order time derivatives proper and the expression for the current in terms of fractional-order derivatives. However, as the example of a comb structure shows, fractional-order time derivatives can only arise in the study of dissipative diffusion processes, and the different equations describe different physical situations. The existence of dead ends alters the order of the time derivative, while the capture of particles by traps changes in the final analysis the expression for the diffusion current.

The author is grateful to Prof. D. E. Khmel'nitskiĭ for support of this work.

*)E-mail: varkhin@bsc.buriatia.ru

[1] Y. Gefen, A. Aharony *et al.*, Phys. Rev. Lett. **47**, 1771 (1981).
[2] R. Alexander and R. Orbach, J. Phys. (France) Lett. **43**, L625 (1982).
[3] R. B. Pandey and D. Stauffer, Phys. Rev. Lett. **51**, 527 (1983).
[4] Y. Gefen, A. Aharony, and S. Alexander, Phys. Rev. Lett. **50**, 77 (1983).
[5] G. Weiss and S. Havlin, Physica A **134**, 474 (1986).
[6] V. E. Arkhincheev and É. M. Baskin, Zh. Éksp. Teor. Fiz. **100**, 292 (1991) [Sov. Phys. JETP **73**, 161 (1991)].
[7] V. E. Arkhincheev, AMSE Press: Modeling, Measurement and Control **49**, 11 (1993).
[8] M. M. Dzhrbashyan, *Integral Transformations and Representations of Functions in the Complex Plane* [in Russian], Nauka, Moscow (1966).
[9] E. Montroll and G. Weiss, J. Math. Phys. **6**, 167 (1965).
[10] M. Shlesinger, J. Stat. Phys. **10**, 421 (1974).
[11] P. P. Nigmatullin, Teoret. Mat. Fiz. **90**, 354 (1992) [Teor. Mat. Fiz. **90**, 242 (1992)].
[12] A. I. Olemskoĭ and A. Ya. Flat, Usp. Fiz. Nauk **163**(12), 1 (1993) [Phys. Usp. **36**, 1087 (1993)].

Translated by Eugene Yankovsky

Photoinduced superstructure in the low-temperature phase of a Peierls system

A. L. Semenov*)

Ul'yanovsk State University, 432700 Ul'yanovsk, Russia
(Submitted 20 June 1998)
Zh. Éksp. Teor. Fiz. **115**, 1297–1314 (April 1999)

This paper is a theoretical study of the properties of the low-temperature phase of a Peierls system when nonequilibrium electron–hole pairs are excited in the phase. A microscopic theory is developed to show that at low temperatures a spatially nonuniform periodic structure with a modulated band gap forms in the thermodynamically nonequilibrium system considered. The critical temperature of formation of such a superstructure, the critical electron–hole pair concentration, the spatial period, and the percentage modulation are calculated. © *1999 American Institute of Physics.* [S1063-7761(99)01004-5]

1. INTRODUCTION

At low temperatures, a one-dimensional chain of equidistantly positioned atoms each of which contains one outer electron is known to become unstable against crystal-lattice period doubling.[1] The corresponding phase transition, characterized by structural distortions and by formation of a band gap in the electron spectrum, is known as the Peierls transition, and the system where such a transition occurs a Peierls system.

Theoretical results obtained from the Peierls model are used to describe the experimentally observed properties of many quasi-one-dimensional materials.[1–6] Among the best-studied materials is vanadium dioxide, whose one-dimensional electron conduction band forms because of the overlap of the $3d$-wave functions of the vanadium atoms, which appear as chains parallel to the crystallographic axis **C** (Ref. 4). Below 340 K, vanadium atoms in a chain converge pairwise and a forbidden band forms in the electron spectrum at the Fermi level, so that the low-temperature phase of VO_2 can be assumed to be a one-dimensional (quasi-one-dimensional) Peierls semiconductor.[6]

Assuming that the Peierls mechanism is the driving force behind the metal–semiconductor phase transition at 340 K in vanadium dioxide and that Hubbard repulsion among the electrons at a single site is negligible, a group of researchers[7–13] theoretically investigated and interpreted a large body of experimental data related to studies of the effect on this transition of various external factors, such as uniaxial and hydrostatic uniaxial pressure,[7] alloying by substitutional impurities,[8–11] the interaction of a vanadium dioxide film and the substrate,[12] and adsorption.[13]

It is of interest to study the behavior of the low-temperature phase of a Peierls system when nonequilibrium electron–hole pairs are excited in the phase. To the author's knowledge, this problem was first examined by Berggren and Huberman,[14] who used numerical analysis to show that, due to strong electron–phonon coupling, a rise in the electron–hole pair concentration leads to a narrowing of the band hap, and the process may be sudden for high levels of excitation. This result has been corroborated by the theory of photoinduced phase transitions in systems with Peierls's instability[15,16] and by data from experiments in which vanadium dioxide films were irradiated by high-power laser pulses.[17–19]

Berggren and Huberman[14] used the phenomenological Ginzburg–Landau expansion of the free energy in powers of the order parameter of the phase transition to show that the uniform semiconducting phase of a Peierls system at low temperatures and high concentrations of nonequilibrium (e.g., photoinduced) electron–hole pairs is unstable against the formation of a periodic superstructure with a spatially modulated band gap. The instability is due to the strong dependence of the electron spectrum on the electron concentration in the conduction band.[15] It is known that an instability of a similar type can also be caused by a strong dependence of the band gap on temperature or deformation of the crystal lattice,[20] or by variations in the dielectric constant generated by band-gap variations.[15] The phenomenological diffusion–deformation–drift models of instability discussed in Refs. 14, 15, and 20 describe the time-dependent regime in the initial stages of superstructure formation.

Note that spatially and temporally nonuniform solutions were analyzed by Mamin[21] and Kopaev *et al.*,[22] who pointed out the possibility of emergence of moving superstructures and solitons[21] and of time-periodic variations in the band gap in the electron spectrum of a system with Peierls's instability.[22]

This paper develops a microscopic theory of steady-state (i.e., already formed) superstructure. The theory is based on a generalization, to the case of nonequilibrium systems, of the mechanism of low-temperature instability of a crystal lattice against static distortions with a wave vector **q**, when the electron spectrum $\varepsilon(\mathbf{k})$ satisfies the nesting condition[23]

$$\varepsilon(\mathbf{k}) = -\varepsilon(\mathbf{k}+\mathbf{q}) \tag{1.1}$$

for all vectors **k** near the Fermi surface and for a fixed vector **q** lying on the Fermi surface.

In the semiconducting phase of the Peierls system and at high levels of excitation of nonequilibrium electron–hole pairs, the Fermi quasilevels of the valence and conduction bands lie in the respective allowed bands of the electron

spectrum. In view of the one-dimensional nature of the system and the symmetry of $\varepsilon(\mathbf{k})$, the condition (1.1) is met near each quasilevel, which results in a transformation of crystal lattice at low temperatures. Here secondary forbidden bands form near the Fermi quasilevels, which in the case at hand is equivalent to spatial modulation of the order parameter of the metal–semiconductor phase transition of the closely related band gap in the electron spectrum of the system.

2. SYSTEM HAMILTONIAN

Let us examine a chain of atoms each of which has one outer electron. The Hamiltonian of the electron subsystem in the tight binding approximation can be written as[1]

$$H=\sum_m B_{m,m+1}(a_m^+ a_{m+1}+a_{m+1}^+ a_m), \tag{2.1}$$

where m is the number of the atom in the chain, $B_{m,m+1}$ is the overlap integral of the wave functions of neighboring electrons, and a_m^+ and a_m are the operators of electron creation and annihilation at the mth atom.

For narrow-gap systems, e.g., for the Peierls model, the separation of adjacent atoms, $r_{m,m+1}$, exceeds the effective radius R of the atomic wave function of an electron several-fold. In this case the overlap integral $B_{m,m+1}$ is given by the expression[24]

$$B_{m,m+1}\propto \exp\left(-\frac{r_{m,m+1}}{R}\right). \tag{2.2}$$

We write the coordinate of the mth site in the chain with spatially modulated pairwise convergence of atoms as follows:

$$x_m=mr_0+\frac{R\xi}{2}\cos(\pi m)\left\{1+\zeta\cos\left[k_0\left(m-\frac{1}{2}\right)\right]\right\}, \tag{2.3}$$

where r_0 is the atomic separation in the metallic phase; ξ is the period-doubling parameter for a one-dimensional crystal, which characterizes the pairwise convergence of atoms (the order parameter of the metal–semiconductor phase transition); ζ is the parameter of modulation of ξ with a wave vector $k_0=2\pi/j$, with j the number of atoms in the chain over one spatial period of the superstructure.

At $\zeta=0$, formula (2.3) describes spatially uniform pairwise convergence of atoms, which characterizes the change in the structure of the lattice under a metal–semiconductor phase transition in the Peierls system.[1] When $\zeta\neq 0$, there is spatial modulation of the structural distortions of the one-dimensional crystal, which leads to similar modulation of the band gap in the semiconducting phase of the Peierls system.

Formula (2.3) is written in such a way that in the Fourier spectrum of the static displacements of the atoms from the equidistant positions of equilibrium there are three modes with the wave numbers $q=\pi$, $q=(\pi-k_0)$, and $q=(\pi+k_0)$. This leads, as we will shortly see (Secs. 3 and 4), to the formation in the electron spectrum of the Hamiltonian (2.1) of forbidden bands at points where the electron quasi-wave number is $k=\pm\pi/2$, $k=\pm(\pi-k_0)/2$, and $k=\pm(\pi+k_0)/2$ (see Fig. 1). If the Fermi level (or quasilevel) is in a forbidden band, the given transformation of the electron spectrum reduces the free energy of the electron subsystem and under certain conditions may correspond to a new stable state of dynamic equilibrium [see the emergence of stable nontrivial solutions $\xi\neq 0$, $\zeta\neq 0$ in Eqs. (5.4) and (5.5)]. The term 1/2 in the cosine in (2.3) has no effect on the final result and has been introduced to simplify (2.4) and all subsequent formulas.

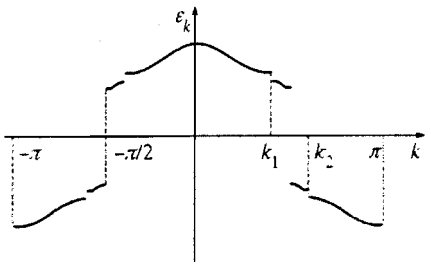

FIG. 1. Sketch of the k-dependence of the electron spectrum ε_k [Eqs. (4.4) and (4.5)] of the Hamiltonian specified by (2.5) and (2.4) of the superstructure in the Peierls system with pairwise convergence of atoms [Eq. (2.3)].

With allowance for the fact that $r_{m,m+1}=x_{m+1}-x_m$, Eqs. (2.2) and (2.3) in the approximation $\zeta\ll 1$ yield

$$B_{m,m+1}=B_m^{(1)}+B_m^{(2)}=b\,\exp[(-1)^m\xi] + b\xi\zeta\,\cos(\pi m)\,\cos(k_0 m)\,\cos(k_0/2), \tag{2.4}$$

where b is the overlap integral in the metallic phase (at $\xi=0$). With the Hamiltonian of the electron subsystem written in the form (2.1), the phases of the wave functions are selected so that b in (2.4) is a real quantity.

Substituting (2.4) into (2.1), we finally obtain

$$H=\sum_{i=1}^{2} H_i, \quad H_i=\sum_m B_m^{(i)}(a_m^+ a_{m+1}+a_{m+1}^+ a_m). \tag{2.5}$$

Note that in view of the approximation $\zeta\ll 1$, adopted in the derivation of (2.4), $H_2\ll H_1$ in the Hamiltonian (2.5).

3. ELECTRON SPECTRUM OF THE SPATIALLY UNIFORM SYSTEM

We begin with a spatially uniform Peierls system ($\zeta=0$). In (2.5) we have $H=H_1$. To diagonalize the Hamiltonian (2.5) we employ Bogolyubov's method of canonical transformations.[25] We introduce the collective second-quantization Fermi operators c_k and c_k^+ as follows:

$$a_m=\frac{1}{\sqrt{N}}\sum_k c_k\, e^{ikm}, \tag{3.1}$$

where N is the number of atoms in the chain, $k=0$, $\pm 2\pi/N$, $\ldots$, $\pm\pi$, and $c_{k+2\pi}=c_k$. In the new operator representation the Hamiltonian (2.5) becomes

$$H_1=\sum_k 2b(c_k^+ c_k\cosh\xi\cos k+ic_k^+ c_{k-\pi}\sinh\xi\sin k). \tag{3.2}$$

In (3.2) we apply another canonical transformation to the operators α_k and α_k^+:

$$c_k=\frac{\alpha_k+i\varphi_k\alpha_{k-\pi}}{\sqrt{1+\varphi_k^2}}. \tag{3.3}$$

The function φ_k in (3.3) is selected in such a way that the resulting Hamiltonian is diagonal in the new variables α_k and α_k^+:

$$H_1=\sum_k E_k\alpha_k^+\alpha_k. \tag{3.4}$$

Substituting (3.3) into (3.2) and zeroing out the off-diagonal elements, we obtain an expression for φ_k and a dispersion law for E_k:

$$\varphi_k=\frac{\cosh\xi\cos k-\mathrm{sgn}(\cos k)\sqrt{\cos^2 k+\sinh^2\xi}}{\sinh\xi\sin k}, \tag{3.5}$$

$$E_k=2b\ \mathrm{sgn}(\cos k)\sqrt{\cos^2 k+\sinh^2\xi}. \tag{3.6}$$

We see that for $\xi\neq 0$ the spectrum E_k has two bands, with the lower band in the ground state completely occupied and the upper band vacant (the semiconducting phase). At $\xi=0$ the spectrum (3.6) consists of one half-filled band (the metallic phase).

4. ELECTRON SPECTRUM OF THE SPATIALLY NONUNIFORM SYSTEM

We now turn to the case where $\zeta=0$ in (2.3)–(2.5). To calculate the electron spectrum of the Hamiltonian (2.5), we use the perturbation-theory approach,[26] bearing in mind that $H_2\ll H_1$. The matrix elements of the perturbation operator H_2 [Eqs. (2.5) and (2.4)] in the representation of the second-quantization Fermi operators c_k [Eq. (3.1)] have the form

$$(H_2)_{pk}=d_k\delta_{p,k+k_0+\pi}+h_k\delta_{p,k-k_0+\pi}, \tag{4.1}$$

where

$$\begin{aligned} d_k&=ib\xi\zeta\exp\left(-\frac{ik_0}{2}\right)\sin\left(k+\frac{k_0}{2}\right)\cos\frac{k_0}{2},\\ h_k&=ib\xi\zeta\exp\left(\frac{ik_0}{2}\right)\sin\left(k-\frac{k_0}{2}\right)\cos\frac{k_0}{2}. \end{aligned} \tag{4.2}$$

When we pass to the Fermi operators α_k of (3.3), the matrix elements (4.1) of the Hamiltonian H_2 of (2.5) become

$$\begin{aligned} (H_2)_{sq}=&[(1+\varphi_s^2)(1+\varphi_q^2)]^{-1/2}\\ &\times[(d_q\delta_{s,q+k_0+\pi}+h_q\delta_{s,q-k_0+\pi})(1-\varphi_q\varphi_s)\\ &-i(d_q\delta_{s,q+k_0}+h_q\delta_{s,q-k_0})(\varphi_q+\varphi_s)]. \end{aligned} \tag{4.3}$$

Bearing this in mind, we can find an approximate expression for the electron spectrum ε_k of the Hamiltonian (2.5) in second-order perturbation theory for $k\in[0,\pi/2]$:

$$\varepsilon_k=\tfrac{1}{2}[E_k+E_q+\mathrm{sgn}(E_k-E_q)\sqrt{(E_k-E_q)^2+4|H_k|^2}], \tag{4.4}$$

where

$$q=k+k_0+\pi,\quad H_k=\frac{\xi\zeta b(1-\varphi_k\varphi_{k+k_0})\sin(k+k_0/2)}{(1+\varphi_k^2)(1+\varphi_{k+k_0}^2)}, \tag{4.5}$$

and E_k is the electron spectrum of the unperturbed system (3.6). For values of the quasiwave number $k\in[-\pi,0]\cup[\pi/2,\pi]$, the spectrum ε_k can be found from (4.4) and (4.5) with allowance for parity and symmetry, i.e., $\varepsilon_k=\varepsilon_{-k}$ and $\varepsilon_k=-\varepsilon_{k+\pi}$. A sketch of ε as a function of k is depicted in Fig. 1.

We see that the pairwise convergence of atoms leads to the formation of forbidden bands in the electron spectrum at points $k\mp\pi/2$ (see Eq. (3.6)), while spatial modulation (2.3) of this pairwise convergence leads to the formation of secondary forbidden bands at points $k=\mp k_1,\mp k_2$, where $k_{1,2}=(\pi\mp k_0)/2$.

Let us study more thoroughly the k-dependence of ε_k for $k\in[0,\pi/2]$. Bearing in mind that the electron spectrum (4.4) differs substantially from (3.6) only in the region $0\leqslant\pi/2-k\leqslant k_0\ll 1$, if we use (3.5) in (4.5) for H_k, we can write the approximate expressions

$$H_k\simeq\begin{cases} b\xi\zeta, & k\in(\pi/2-k_0,\pi/2),\\ 0, & k\in[0,\pi/2-k_0]. \end{cases} \tag{4.6}$$

Substituting (4.6) into (4.4) and keeping only the quadratic approximation for E_k near the bottom of the conduction band, we finally obtain

$$\varepsilon_k\simeq\begin{cases} E_{k_1}+\dfrac{b(k-k_1)^2}{\sinh\xi}-b\,\mathrm{sgn}(k-k_1)\\ \times\sqrt{\left[\dfrac{k_0(k-k_1)}{\sinh\xi}\right]^2+(\xi\zeta)^2}, & k\in[\pi/2-k_0,\pi/2),\\ E_k, & k\in[0,\pi/2-k_0]. \end{cases} \tag{4.7}$$

Thus, as Eqs. (4.7) imply, when there is spatial modulation (2.3) at point $k=k_1=(\pi-k_0)/2$ there forms a forbidden band, or gap, $\Delta\varepsilon_g=2b\xi\zeta$.

Note that the form of (4.3) suggests that a forbidden band forms also at $k=k_0/2$. Here, ε_k near $k=k_0/2$ has the form (4.4), where now $q=k-k_0$ and

$$H_k\simeq b\left(k-\frac{k_0}{2}\right)^2\xi\zeta\tanh\xi. \tag{4.8}$$

Equation (4.8) shows, however, that $H_{k_0/2}=0$, and hence at $k=k_0/2$ no forbidden bands form in the electron spectrum.

5. EQUILIBRIUM EQUATIONS

Let us examine the behavior of the low-temperature phase of a Peierls system when nonequilibrium electron–hole pairs are excited in the phase. We assume that the process is due to stimulated transitions of electrons from the valence band to the conduction band that occur, for instance, because of the dipole electron–photon interaction with the incident radiation. As is known, the characteristic intraband relaxation time of electrons, $\tau_e\sim 10^{-14}$ s, is much shorter than the interband relaxation time $\tau\sim 10^{-11}$ s (see Ref. 27). Hence, when a light field with a constant amplitude irradiates

the system, we can approximately assume[28] that within each electron band there is thermodynamic equilibrium between the electrons, with a Fermi quasilevel corresponding to each band. Violation of the thermodynamic equilibrium between the bands caused by external light manifests itself in the difference between the various Fermi quasilevels

The above approach to describing a thermodynamically nonequilibrium system consisting of a set of thermodynamically equilibrium subsystems can be generalized to the case where the incident radiation has an adiabatically slowly-varying amplitude A (the variation ΔA of the amplitude over a time interval $\tau_e \sim 10^{-14}$ s is much smaller than A). The reason for this is that the electron subsystem within each band has time to closely follow the field variations, so that at each moment the subsystem is in thermodynamic equilibrium. To a certain extent this situation is similar to the one usually encountered in the description of thermodynamically equilibrium systems when the external parameters vary adiabatically slowly.

Below we limit ourselves to building a theory for this specific case. Transient processes that take $\tau_e \sim 10^{-14}$ s when a steep leading or trailing edge of the light pulse passes through the system will not be discussed.

The free energy F_j of the electron subsystem of the jth band ($j=1,2$) is specified by

$$F_j = \mu_j N_j - k_B T \sum_k \ln\left[1+\exp\left(\frac{\mu_j-\varepsilon_k}{k_B T}\right)\right], \tag{5.1}$$

where μ_j and N_j are, respectively, the Fermi quasilevel and the number of electrons of the jth band. Summation over k in (5.1) is done within the limits of the jth band specified by (4.4) ($|k|<\pi/2$ at $j=1$ and $\pi/2<|k|<\pi$ at $j=2$).

As is known, the characteristic relaxation time of the phonon subsystem is $\tau_{\rm ph} \sim 10^{-13}$ s (see Ref. 27). Hence at the moments when the amplitude of the pulse of incident radiation changes insignificantly during $\tau_{\rm ph} \sim 10^{-13}$ s (adiabatically slow variation of the external parameter), the phonon subsystem has time to relax to its dynamically equilibrium state, which depends on the instantaneous amplitude of the pulse.

In this case the expression for free energy of the crystal lattice with allowance for structural distortions [Eq. (2.3)] can be written in the harmonic approximation as follows:

$$F_c = F_0 + \frac{\gamma}{2}\sum_m (r_{m,m+1}-r_0)^2, \tag{5.2}$$

where F_0 is the free energy (it characterizes the dynamics of the lattice), and γ is the stiffness of the lattice under static displacements (2.3) of the atoms. This formula is written in the molecular field approximation,[1] in which it is assumed that the phonon part F_0 does not depend on the parameters ξ and ζ, which are responsible for static distortions. From a physical standpoint the given approximation means that there is no interaction between dynamic ($\omega \neq 0$) and static ($\omega=0$) phonon modes.

If we combine (2.3) and the fact that $r_{m,m+1}=x_{m+1}-x_m$ with (5.2), we obtain

$$F_c = F_0 + \frac{A}{2}\xi^2\left[1+\frac{1}{2}\zeta^2\cos^2\left(\frac{k_0}{2}\right)\right], \tag{5.3}$$

where $A=\gamma N R^2$.

For a thermodynamically nonequilibrium system to be in the steady state of dynamic equilibrium, the generalized forces f_1 and f_2 corresponding to the generalized coordinates ξ and ζ must vanish:

$$f_1 \equiv -\left(\frac{\partial F_1}{\partial \xi}\right)_{T,N_1} - \left(\frac{\partial F_2}{\partial \xi}\right)_{T,N_2} - \left(\frac{\partial F_c}{\partial \xi}\right)_T = 0, \tag{5.4}$$

$$f_2 \equiv -\left(\frac{\partial F_1}{\partial \zeta}\right)_{T,N_1} - \left(\frac{\partial F_2}{\partial \zeta}\right)_{T,N_2} - \left(\frac{\partial F_c}{\partial \zeta}\right)_T = 0. \tag{5.5}$$

If we combine these equations with (5.1) and (5.3), we find that

$$f_1 = -A\xi\left[1+\frac{1}{2}\zeta^2\cos^2\left(\frac{k_0}{2}\right)\right] + 2\sum_{|k|\leqslant\pi/2}\frac{\partial\varepsilon_k}{\partial\xi}\tanh\left(\frac{\varepsilon_k-\mu}{2k_BT}\right)=0, \tag{5.6}$$

$$f_2 = -\frac{1}{2}A\xi^2\zeta\cos^2\left(\frac{k_0}{2}\right) + 2\sum_{|k|\leqslant\pi/2}\frac{\partial\varepsilon_k}{\partial\zeta}\tanh\left(\frac{\varepsilon_k-\mu}{2k_BT}\right)=0. \tag{5.7}$$

The expressions (5.6) and (5.7) are the equations of equilibrium of a Peierls system. They determine the behavior of the parameters ξ and ζ when nonequilibrium electron–hole pairs are excited.

Physically, it is convenient to adopt the total concentration n of electron–hole pairs (which includes equilibrium and nonequilibrium excitations) as the external control parameter characterizing the effect of light on the system. To do this, we must write, in addition to (5.6) and (5.7), an equation that reflects the electroneutrality of the system (this equation links n and the Fermi quasilevel μ):

$$n = \frac{N}{2} - \sum_{|k|\leqslant\pi/2}\tanh\frac{\varepsilon_k-\mu}{2k_BT}. \tag{5.8}$$

Thus, Eqs. (5.6), (5.7), and (5.8) form a complete set of equations for determining the parameters ξ and ζ of structural distortions of the lattice [Eq. (2.3)] at a given temperature T and a given electron–hole pair concentration n.

6. FORMATION OF A PERIODIC SUPERSTRUCTURE AT ABSOLUTE ZERO *T=0*

We begin our analysis of Eqs. (5.6)–(5.8) with the case $T=0$ under the assumption that the superstructure modulation parameter ζ is much smaller than unity. Then Eq. (5.6) yields

$$A\xi - \frac{4bN}{\pi}\left(K\left(\sqrt{1-\tanh^2\xi}\right)\sinh\xi - 2\xi\ln\times\left|\frac{k_0/2+\sqrt{(k_0/2)^2+\xi^2}}{\xi}\right|\right)=0, \tag{6.1}$$

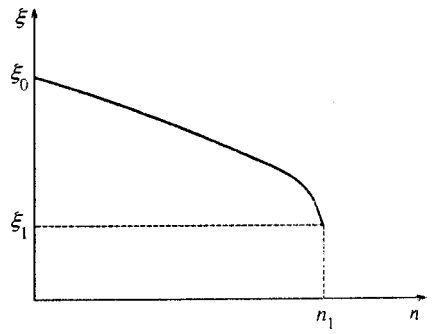

FIG. 2. Sketch of the dependence of the order parameter ξ of the metal–semiconductor phase transition on the total concentration n of electron–hole pairs. At $T=0$ the quantities ξ_1 and n_1 are given by (6.6), while for $T\neq 0$ [with allowance for the condition (7.2)] they are given by (7.5).

where $K(x)$ is the complete normal elliptic integral of the first kind.

Equation (6.1) describes the relationship between the parameter ξ and the concentration n of the nonequilibrium electron–hole pairs in a Peierls system. If we allow for the fact that in real physical systems the concentration n of electron–hole pairs is much lower than N and that the order parameter ξ of the metal–semiconductor phase transition does not exceed 0.5 (see Refs. 1–6), Eq. (6.1) yields the approximate relationship

$$n=\frac{N}{\pi}(\xi_0-\xi)\sqrt{\frac{\xi}{\xi_0}}, \tag{6.2}$$

where

$$\xi_0=\frac{\pi}{2}\exp\left\{\sin^{-1}\left(\frac{\pi}{4}\right)-\frac{A\pi}{4bN}\right\} \tag{6.3}$$

is order parameter of the metal–semiconductor phase transition at $n=0$.

In deriving Eqs. (6.2) and (6.3) we also used the equation

$$k_0=\frac{\pi n}{N}, \tag{6.4}$$

which is valid at $T=0$. A rough sketch of the ξ vs. n dependence with allowance for the condition of stability under spatially uniform fluctuations of the order parameter ξ of the metal–semiconductor phase transition,

$$\frac{\partial n}{\partial \xi}<0, \tag{6.5}$$

is depicted in Fig. 2. We see that as the electron–hole pair concentration n increases to the value n_1, the order parameter ξ of the metal–semiconductor phase transition smoothly decreases to ξ_1, where

$$n_1=\frac{2N\xi_0}{3\sqrt{3}\,\pi}, \quad \xi_1=\frac{\xi_0}{3}. \tag{6.6}$$

At point n_1 the value of ξ suddenly drops from ξ_1 to zero (the phase transition to the metallic state.)

Combing Eq. (5.7) with (4.7) and (5.8), we obtain an expression for the superstructure modulation parameter ξ at $T=0$:

$$\zeta=\frac{k_0^2}{2\xi\sinh\xi}\sinh^{-1}\frac{\pi A k_0\cos^2(k_0/2)}{16bN\sinh\xi}. \tag{6.7}$$

Taking (6.4) into account and assuming that $k_0=\pi n/N\ll 1$, we see that (6.7) yields the approximate formula

$$\zeta=\frac{8bn}{A\xi}. \tag{6.8}$$

Thus, at $T=0$ a superstructure in a Peierls system exists in the entire semiconducting phase for $n\in(0,n_1)$. Here the modulation parameter ζ given by (6.8) monotonically increases with n from the value $\zeta(n=0)=0$ to the value $\zeta 1$:

$$\zeta_1=\zeta(n=n_1)=\frac{16bN}{\pi\sqrt{3}A}, \tag{6.9}$$

while the spatial period of the system,

$$\lambda=\frac{2\pi r_0}{k_0}=\frac{2r_0N}{n}, \tag{6.10}$$

monotonically decreases from $\lambda(n=0)=\infty$ to

$$\lambda_1=\lambda(n=n_1)=\frac{3\sqrt{3}\,\pi r_0}{\xi_0}. \tag{6.11}$$

Here are some numerical estimates. To calculate the stiffness coefficient A of the crystal lattice of the Peierls system, we use the formula[10]

$$A=\frac{4bN}{\pi}\left[\ln\left(\frac{\pi b}{2k_BT_0}\right)+1\right], \tag{6.12}$$

where T_0 is the critical temperature of a thermodynamically equilibrium metal–semiconductor transition in the Peierls system.

If we use typical numerical values of the physical quantities for VO_2 (see Refs. 4 and 7), i.e., $b\approx 0.3\,\text{eV}$, $T_0\approx 340\,\text{K}$, $r_0\approx 3\,\text{Å}$, $N\approx 10^{23}\,\text{cm}^{-3}$, and $\xi_0\approx 0.5$, then Eqs. (6.12), (6.6), (6.9), and (6.11) suggest that $A\approx 10^{23}\,\text{eV/cm}^3$, $n_1\approx 6\times 10^{21}\,\text{cm}^{-3}$, $\zeta_1\approx 0.9$, and $\lambda_1\approx 100\,\text{Å}$.

7. FORMATION OF A PERIODIC SUPERSTRUCTURE AT $T\neq 0$

We analyze Eqs. (5.6)–(5.8) with $T\neq 0$ under the assumption that the Peierls system is a nondegenerate or weakly degenerate semiconductor:

$$\mu-2b\sinh\xi<2k_BT. \tag{7.1}$$

This relationship, which imposes a restriction on the size of the region within which the Fermi quasilevel μ may vary, is equivalent, if we allow for Eq. (5.8), to an approximate inequality, an upper bound on the concentration n of the electron–hole pairs:

$$n<n_2=\frac{8N}{3\pi}\sqrt{\frac{k_BT\sinh\xi}{b}}. \tag{7.2}$$

Combining (5.6) and (5.8) and assuming that the superstructure modulation parameter ζ is much smaller than unity, we obtain an approximate equation for the order parameter ξ of the metal–semiconductor phase transition:

$$A\xi-4b\left[\frac{N}{\pi}K(\sqrt{1-\tanh^2\xi})\ \sinh\ \xi-n\ \cosh\ \xi\right]=0, \qquad (7.3)$$

where $K(x)$ is the complete normal elliptic integral of the first kind.

Bearing in mind that in real physical systems the electron–hole pair concentration $n\ll N$ and that the order parameter ξ of the metal–semiconductor phase transition does not exceed 0.5 (see Refs. 1–6), from (7.3) we obtain the approximate equation

$$n=\frac{N\xi}{\pi}\ln\frac{\xi_0}{\xi}, \qquad (7.4)$$

with ξ_0 determined by (6.3).

Figure 2 is a rough sketch of that segment of $\xi(n)$ given by (7.4) where the stability condition (6.5) is satisfied. We see that the behavior of $\xi(n)$ is similar to that in a system with $T=0$, but now the sudden transition to the metallic phase takes place at

$$n_1=\frac{N\xi_0}{\pi e}, \quad \xi_1=\frac{\xi_0}{e}. \qquad (7.5)$$

The inequality (6.5) is the criterion of stability for the semiconducting phase of the Peierls system against a metal–semiconductor phase transition. Now we study this system for stability under a transition to a spatially nonuniform state with a periodic spatial modulation order parameter ξ of the metal–semiconductor phase transition ($\zeta\neq0$). Bearing in that at the moment at which the solution loses its stability, $\zeta=0$,

$$\left.\frac{\partial f_2}{\partial\zeta}\right|_{T,\zeta=0}=0, \qquad (7.6)$$

from Eq. (5.7) we obtain

$$T=\frac{2b^2Nk_0}{\pi Ak_B\ \cos^2\ (k_0/2)}. \qquad (7.7)$$

When a photoinduced superstructure with a wave number k_0 is formed, the forbidden band in the electron spectrum $E(k)$ [see Eq. (3.6)] emerges at the point where the Fermi quasilevel μ is located:

$$E\left(\frac{\pi-k_0}{2}\right)=\mu. \qquad (7.8)$$

This yields the approximate expression Equation (7.9)

$$\mu=2b\ \sinh\ \xi+\frac{bk_0^2}{4\ \sinh\ \xi}. \qquad (7.9)$$

According to Eq. (5.8), when the Fermi quasilevel μ is in the conduction band, the concentration n of electron–hole pairs is given by the approximate expression

$$n=\frac{N}{3\pi k_BT}\sqrt{\frac{\sinh\ \xi}{b}(\mu+2k_BT-2n\ \sinh\ \xi)^3}. \qquad (7.10)$$

If we now use Eqs. (7.7)–(7.10), we can find a formula for the critical concentration n_c of electron–hole pairs above which ($n>n_c$) the semiconducting phase of the Peierls system contains a photoinduced superstructure:

$$n_c=\frac{2N}{3\pi}\sqrt{\frac{2k_BT\ \sinh\ \xi}{b}\left[1+\frac{k_BT}{32b^3\ \sinh\ \xi}\left(\frac{\pi A}{N}\right)^2\right]^3}. \qquad (7.11)$$

Combining Eqs. (7.9) and (7.10), we obtain an expression for the spatial period λ of the superstructure:

$$\lambda=\frac{2\pi r_0}{k_0}$$

$$=\frac{\pi r_0}{\sqrt{(3\pi k_BTn\ \sinh\ \xi/Nb)^{2/3}-2k_BT\ \sinh\ \xi/b}}. \qquad (7.12)$$

In particular, Eqs. (7.11) and (7.12) imply that at the critical point n_c, at the moment when the superstructure is formed, the value of the period $\lambda_c=\lambda(n_c)$ is given by

$$\lambda_c=\frac{4b^2Nr_0}{Ak_BT}. \qquad (7.13)$$

To observe the superstructure, the following conditions must be met:

$$n_c<n<n_1. \qquad (7.14)$$

which imply, in particular, that $n_c<n_1$. If we combine this fact with (7.5) and (7.11), we obtain an approximate equation for the critical temperature T_c above which there can be no superstructure, no matter what the value of n is:

$$\frac{8k_BT_c}{9b\xi_0}\left[1+\frac{ek_BT_c}{32b^3\xi_0}\left(\frac{\pi A}{N}\right)^2\right]^3=1. \qquad (7.15)$$

Let us now give the results of numerical estimates. If we use typical numerical values of the physical quantities for VO_2 (Refs. 4 and 7), i.e., $b\approx0.3$ eV, $r_0\approx3$ Å, $N\approx10^{23}\,\mathrm{cm}^{-3}$, $\xi_0\approx0.5$, $A\approx10^{23}\,\mathrm{eV\,cm}^3$, and $T\approx100$ K, then Eqs. (7.2), (7.5), (7.11), (7.13), and (7.15) suggest that $n_2\approx10^{22}\,\mathrm{cm}^{-3}$, $n_1\approx6\times10^{21}\,\mathrm{cm}^{-3}$, $n_c\approx5\times10^{21}\,\mathrm{cm}^{-3}$, $\lambda_c\approx130$ Å, and $T_c\approx200$ K. Note that here the numerical value of n_1 coincides, in order of magnitude, with the value estimated from the experimental data, $n_1\approx10^{21}\,\mathrm{cm}^{-3}$ (see Ref. 18).

8. DIPOLE-MOMENT OPERATOR

From now on we will assume that the nonequilibrium concentration of the electron–hole pairs in the system considered is created thanks to the electric dipole electron–photon interaction with the incident radiation. To describe this interaction, we first calculate the dipole-moment operator of a spatially uniform ($\zeta=0$) Peierls system. In the tight binding approximation, this dipole-moment operator is

$$\mathbf{d}=\sum_n\ (\mathbf{d}_{n,n+1}a_n^+a_{n+1}+\mathbf{d}_{n,n+1}^*a_{n+1}^+a_n), \qquad (8.1)$$

where the dependence of $\mathbf{d}_{n,n+1}$ on ξ is similar to (2.4):

$$\mathbf{d}_{n,n+1}=(\mathbf{d}_1+i\mathbf{d}_2)\exp((-1)^n\xi)$$
$$=-e\int \psi_n^*(\mathbf{r})\mathbf{r}\psi_{n+1}(\mathbf{r})\,d\mathbf{r}. \tag{8.2}$$

Here $\psi_n(\mathbf{r})$ is the atomic wave function of the electron at the nth site, and e is the electron charge. By selecting the phases of the wave functions $\psi_n(\mathbf{r})$ so that the overlap integral (2.4) is real we ensure that both $\mathbf{d}_1$ and $\mathbf{d}_2$ in (8.2) are uniquely defined.

Substituting (3.1) into (8.1) and allowing for (8.2) and the identity $\exp[(-1)^n\xi]=\cosh\xi+(-1)^n\sinh\xi$, we obtain

$$\mathbf{d}=2\sum_k [\cosh\xi(\mathbf{d}_1\cos k-\mathbf{d}_2\sin k)c_k^+c_k$$
$$+i\sinh\xi(\mathbf{d}_1\sin k+\mathbf{d}_2\cos k)c_k^+c_{k-\pi}]. \tag{8.3}$$

Introducing the Fermi operators α_k and α_k^+ into (8.3) and allowing for (3.3) and (3.5) we finally obtain

$$\mathbf{d}=\sum_k \left\{\left[\frac{\mathbf{d}_1E_k}{b}-\frac{2\mathbf{d}_2}{1+\varphi_k^2}[(1-\varphi_k^2)\cosh\xi\sin k\right.\right.$$
$$\left.+2\varphi_k\sinh\xi\cos k]\right]\alpha_k^+\alpha_k+i\frac{2\mathbf{d}_2}{1+\varphi_k^2}[(1-\varphi_k^2)\sinh\xi$$
$$\left.\times\cos k-2\varphi_k\cosh\xi\sin k]\alpha_k^+\alpha_{k-\pi}\right\}. \tag{8.4}$$

Note the formal similarity of the operators (2.1) and (8.1) at $\mathbf{d}_2=0$. Hence the operators (3.4) and (8.4) are also formally similar.

Suppose that without an external electric field the total dipole moment of the system is zero. Then Eqs. (8.4) and (3.6) imply that $\mathbf{d}_1=0$. Thus, the fact that we have chosen the phases of the wave functions $\psi_n(\mathbf{r})$ so that the overlap integral $B_{n,n+1}$ in (2.1) is real ensures, in the present case, that the interstitial dipole-moment matrix element $\mathbf{d}_{n,n+1}$ in (8.1) is imaginary. The case of $\mathbf{d}_1\neq 0$ can probably be observed in systems exhibiting ferroelectric properties, but we will not consider such systems here. As $\xi\to 0$, as Eq. (3.5) indicates, $\varphi_k\to 0$ for all $k\neq\pm\pi/2$, with the result that $\mathbf{d}_{k,k-\pi}\to 0$ in (8.4) and all dipole transitions are forbidden. If $\xi\neq 0$, then $\mathbf{d}_{k,k-\pi}\neq 0$ in (8.4), and the corresponding dipole transitions are allowed. Since in this case the interval $k\in(-\pi/2,\pi/2]$ is the first Brillouin zone, the given transitions in the spectrum (3.6) are vertical band-to-band transitions.

9. INTERACTION WITH RADIATION

The interaction between the system and the light field is described by an operator V, which in the dipole approximation can be written as

$$V=-\mathbf{d}\cdot\mathbf{E}(t)=\mathbf{d}\cdot\int \mathbf{E}_\omega\, e^{-i\omega t}d\omega, \tag{9.1}$$

where $\mathbf{e}_\omega$ and ω are the amplitude and frequency of a spectral component of the light field.

We assume the incident radiation $\mathbf{E}(t)$ to be a quasimonochromatic time-independent random process linearly polarized along the crystal's axis.[29] As a result, all the spectral components $\mathbf{E}_\omega$ are statistically independent:[29,30]

$$\langle\mathbf{E}_\omega\cdot\mathbf{E}_{\omega_1}\rangle=G(\omega)\,\delta(\omega+\omega_1). \tag{9.2}$$

Here $G(\omega)$ is the spectral density of the light field, which for a quasimonochromatic signal can be written as[29]

$$G(\omega)=Ig(|\omega|-\omega_0), \tag{9.3}$$

where ω_0 is the carrier frequency, and $g(x)$ is a nonnegative bell-shaped function, with its maximum at $x=0$, satisfying the normalization $\int g(x)\,dx=1$. The width $\Delta\omega$ of the spectrum $G(\omega)$ satisfies the inequality $\Delta\omega\ll\omega_0$. The quantity $I=\int G(\omega)d\omega/2$ is the intensity of the light field (in the Gaussian system of units, to within a factor $cn/2\pi$, where c is the speed of light, and n is the medium's refractive index).

Using Liouville's equation[31]

$$i\hbar\frac{\partial\rho}{\partial t}=[H+V,\rho], \tag{9.4}$$

and allowing for Eqs. (9.1) and (9.2), we obtain an equation for the diagonal elements ρ_{kk} of the density matrix ρ of the electron subsystem in second-order perturbation theory:

$$\frac{\partial\rho_{kk}}{\partial t}=\frac{2\pi}{\hbar^2}\sum_s |\mathbf{d}_{ks}|^2G\left(\frac{E_s-E_k}{\hbar}\right)(\rho_{ss}-\rho_{kk}), \tag{9.5}$$

where $\mathbf{d}_{ks}$ is the matrix element of the dipole moment operator (8.4). In the special case of a monochromatic field $\mathbf{E}(t)=\mathbf{E}_0\cos(\omega_0t+\varphi)$ with a uniformly distributed random phase φ, the spectral density is

$$G(\omega)=\tfrac{1}{4}\mathbf{E}_0^2[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)].$$

Then Eq. (9.5) becomes Fermi's Golden Rule for the probability of stimulated transitions:[30]

$$\frac{\partial\rho_{k,k-\pi}}{\partial t}=\frac{\pi}{2\hbar}|\mathbf{E}_0\cdot\mathbf{d}_{k,k-\pi}|^2\delta(2E_k-\hbar\omega_0). \tag{9.6}$$

Here (9.6) we assumed that the lower level (with the quasiwave number $k-\pi$) is completely occupied and the upper level (with the quasiwave number k) is vacant.

If we combine (8.4) and (3.6) with (9.5), we obtain

$$\frac{\partial\rho_{kk}}{\partial t}=\frac{2\pi}{\hbar^2}d_k^2G\left(\frac{2\varepsilon_k}{\hbar}\right)\tanh\frac{\varepsilon_k-\mu}{2k_BT}, \tag{9.7}$$

where

$$d_k=\frac{2d_2}{1+\varphi_k^2}|(1-\varphi_k^2)\sinh\xi\cos k-2\varphi_k\cosh\xi\sin k|. \tag{9.8}$$

In deriving Eq. (9.7) we assumed that the spectral density $G(\omega)$ of the light field is localized near the frequencies $\omega_k=2\varepsilon_k/\hbar$ at which the electron spectrum ε_k of a spatially nonuniform Peierls system [see Eq. (4.7)] coincides with the spectrum E_k of a spatially uniform Peierls system [see Eq. (3.6)], i.e., there is excitation of electron–hole pairs into the depth of the allowed band.

Combining Eqs. (3.6) and (9.7), we obtain a transport equation for the concentration $n=2\Sigma_{|k|\leqslant\pi/2}\rho kk$ of the electron–hole pairs:

$$\frac{\partial n}{\partial t}=\frac{4\pi}{\hbar^2}\sum_{|k|\leqslant\pi/2} d_k^2 \tanh\left(\frac{\varepsilon_k-\mu}{2k_BT}\right) G\left(\frac{2\varepsilon_k}{\hbar}\right)-\frac{n-n_0}{\tau}, \quad (9.9)$$

where n_0 is the electron–hole pair concentration in the absence of a light field, i.e., $n_0=n(\mu=0)$. The last term on the right-hand side of Eq. (9.9) allows for band-to-band relaxation with a relaxation time τ. Equation (9.9) shows that the variation in n is due to the interaction of the electron subsystem and the spectral components of the light field with frequencies $\omega_k=2\varepsilon_k/\hbar$.

Taking (9.3) into account, from Eq. (9.9) in the steady-state regime ($\partial n/\partial t=0$) we finally obtain

$$I=\frac{(n-n_0)\hbar^2}{4\pi\tau}\left[\sum_{|k|\leqslant\pi/2} d_k^2 \tanh\left(\frac{\varepsilon_k-\mu}{2k_BT}\right) g\left(\frac{2\varepsilon_k}{\hbar}-\omega_0\right)\right]^{-1}. \quad (9.10)$$

Equations (5.8), (5.6), and (9.10) constitute a complete set of equations with respect to the internal parameters μ, n, and ξ of a spatially uniform ($\zeta=0$) Peierls system with given external parameters T, I, ω_0, etc.

10. EXCITATION BY A MONOCHROMATIC LIGHT FIELD

For a monochromatic light field, we analyze Eq. (9.10) with allowance for (7.4), which holds at T, where the form factor $g(x)=\delta(x)$. In (9.10) we replace the sum by an integral over dE, bearing in mind that the density of electron states $\nu(E)$ corresponding to the spectrum (3.6) takes the form

$$\nu(E)=\frac{2NE}{\pi\sqrt{(4b^2\cosh^2\xi-E^2)(E^2-4b^2\sinh^2\xi)}} \quad (10.1)$$

and that, according to (9.8), (3.5), and (3.6), the matrix element of the dipole-moment operator, d_k, at $\varepsilon_k=E$ is given by

$$d(E)\equiv d_k(\varepsilon_k=E)=\frac{4bd_2\cosh\xi\sinh\xi}{E}. \quad (10.2)$$

Then at $\tanh[(\hbar\omega_0/2-\mu)/2k_BT]=1$ we have the following I vs. ξ dependence:

$$I=I(\xi)$$
$$=\frac{\hbar^2\omega_0[\xi\ln(\xi_0/\xi)-\pi n_0/N]}{512\pi\tau b^2d_2^2\cosh^2\xi\sinh^2\xi}$$
$$\times\sqrt{[16b^2\cosh^2\xi-(\hbar\omega_0)^2][(\hbar\omega_0)^2-16b^2\sinh^2\xi]}. \quad (10.3)$$

Similar calculations with $T=0$, as Eqs. (9.10), (6.2), (10.1), and (10.2) imply, lead to the expression

$$I=I(\xi)$$
$$=\frac{\hbar^2\omega_0(\xi_0-\xi)\sqrt{\xi/\xi_0}}{512\pi\tau b^2d_2^2\cosh^2\xi\sinh^2\xi}$$
$$\times\sqrt{[16b^2\cosh^2\xi-(\hbar\omega_0)^2][(\hbar\omega_0)^2-16b^2\sinh^2\xi]}. \quad (10.4)$$

Thus, when $T=0$, we have (10.4) for $I(\xi)$, (6.2) for $n(\xi)$, (6.8) for $\zeta(n)$, and (6.10) for $\lambda(n)$. These formulas indirectly specify (in parametric form) the electron–hole pair concentration n, the modulation parameter ζ, and the period λ of the superstructure as functions of the light-field intensity I.

When T is finite, we have (10.3) for $I(\xi)$, (7.4) for $n(\xi)$, and (7.12) for $\lambda(n,\xi)$, which implicitly specify $n(I)$ and $\lambda(I)$.

Here are some numerical estimates. We use typical numerical values of the physical quantities for VO_2: $b\approx0.3$ eV, $N\approx10^{23}\,\mathrm{cm}^{-3}$, and $\xi_0\approx0.5$ (Refs. 4 and 7); $\tau\sim10^{-11}$ s (Ref. 27); and $d_2\sim10^{-18}$ esu (Refs. 30 and 32). Then from (10.3) and (10.4) with allowance for (7.5) and (6.6) it follows that at the photoinduced semiconductor–metal phase transition point, $I_1(T=0)\approx I_1(T=100\,\mathrm{K})\sim10^6$ esu. This corresponds to a light-field intensity $I_1c/2\pi\sim10^{15}$ esu$\sim10^8\,\mathrm{W/cm}^2$. A photoinduced superstructure can be observed when $I\in(I_c,I_1)$, with $I_c(T=0)=0$. When $T=100$ K, Eqs. (7.11), (7.4), and (10.3) yield $I_1-I_c\approx2\times10^5$ esu, which corresponds to a light-field intensity $c(I_1-I_c)/2\pi\approx2\times10^{14}$ esu$\approx2\times10^7\,\mathrm{W/cm}^2$.

11. DISCUSSION

The high values of the critical intensities, namely $I_1c/2\pi\approx I_cc/2\pi\sim10^7-10^8\,\mathrm{W/cm}^2$, obtained via the theory elaborated in this paper, appear to be determined by the fact that observing photoinduced effects of formation of a superstructure and the semiconductor–metal phase transition requires high nonequilibrium electron–hole pair concentrations, $n_c\sim n_1\sim0.05N$ (see the results of numerical estimates at the end of Sec. 7). The value of n_c (and hence of I_c) can be reduced by lowering the temperature [see Eq. (7.11)]. However, numerical estimates show that even at $T=1$ K the values of the critical concentration n_c and critical intensity I_c remain very large: $n_c\approx5\times10^{20}\,\mathrm{cm}^{-3}$ and $I_cc/2\pi\approx7\times10^6\,\mathrm{W/cm}^2$.

To observe the predicted phenomena in experiments, one should use very thin (down to atomic dimension) filamentary (one-dimension) or film (two-dimensional) samples placed inside a material that is transparent to radiation at the laser frequency and acts as a good heat sink (say, superfluid helium).

Another possible way to prevent overheating of the system is to use high-power short-pulse laser light. In this case the formulas of Sec. 7, in particular, the expressions for the concentrations n_1 [Eq. (7.5)] and n_c [Eq. (7.11)] remain valid if the laser pulse length exceeds severalfold the characteristic relaxation times $\tau_{\mathrm{ph}}\sim10^{-13}$ s and $\tau_e\sim10^{-14}$ s of the phonon and interband electron subsystems, respectively.

In view of what has just been said, the experiment conducted by Bugaev *et al.*[18] can be cited as an example. The researchers irradiated a vanadium dioxide film by picosecond laser pulses with an intensity $I \approx 5 \times 10^8$ W/cm^2 and observed a photoinduced semiconductor–metal phase transition. Here, as the numerical estimates of the present paper show, the critical concentration of the nonequilibrium carriers at the phase transition point, n_1/N, was roughly 10^{-2}, and the variation in the sample temperature, ΔT, did not exceed 10 K.

An experiment similar to the one described in Ref. 18 but conducted at low temperatures ($T < T_c$) would probably make it possible to observe photoinduced superstructures. The temperature T of the sample should be maintained as low as possible, since according to (7.11) this reduces the threshold value of the critical nonequilibrium electron–hole pair concentration n_c above which a superstructure of the type described in the present paper appears in a Peierls semiconductor.

12. CONCLUSION

We have studied the behavior of the low-temperature phase of a Peierls system when nonequilibrium electron–hole pairs are excited in the phase. We have found that as the total concentration n of the excited electron–hole pairs increases (including thermal and light excitations), the order parameter ξ of the metal–semiconductor phase transition smoothly decreases to the value ξ_1 (Eqs. (7.4) and (7.5) and Fig. 2), and at the concentration n_1 given by Eq. (7.5) a sudden transition to the metallic phase occurs.

At low temperatures $T < T_c$, where T_c is defined in (7.15), and at electron–hole concentrations n such that $n_c < n < n_1$ [see Eqs. (7.11) and (7.5)], a spatially nonuniform periodic structure forms in the Peierls system. This structure is the one-dimensional spatial modulation of the band gap E_p in the electron spectrum with a spatial period $\lambda = 2\pi r_0/k_0$ [see Eq. (7.12)]:

$$E_g = 4b \sinh\left\{\xi\left[1+\zeta\cos\left(k_0\left(m-\frac{1}{2}\right)\right)\right]\right\}, \qquad (12.1)$$

where m is the number of the atom in the chain.

At time of formation, i.e., when $n = n_c$, the period λ of this heterostructure is given by (7.13). Then, as n increases, λ decreases according to Eq. (7.12). After the light pulse has traveled through the system, when the concentration n becomes lower than the critical value ($n < n_c$), the heterostructure disappears.

At $T = 0$, the photoinduced heterostructure exists in the semiconducting phase of the Peierls system at any concentration n lower than n_1 [Eq. (6.6)], since here $n_c = 0$. As n increases, the spatial period λ of the heterostructure [Eq. (6.10)] monotonically decreases from $\lambda(n=0) = \infty$ to $\lambda(n = n) = \lambda_1$ [Eq. (6.11)]. Also, as n increases, the parameter ζ [Eq. (6.8)] of modulation of the band gap [Eq. (12.1)] increases from $\zeta(n=0) = 0$ to $\zeta(n = n_1) = \zeta_1$ [Eq. (6.9)].

Basing our reasoning on the assumption that the photoexcitation of nonequilibrium electron–hole pairs is due to the electric dipole interaction of the electron subsystem and the photons of the incident radiation, we have derived expressions for the value of the light-field intensity I at which observation of the photoinduced superstructure is possible, and made numerical estimates (see Sec. 10). We have found that the photoinduced superstructure forms at intensities $I \in (I_c, I_1)$, where $I_c(T=0) = 0$, and monotonically grows as the intensity increases to $I_c(T = T_c) = I_1$. The value of I_1 changes little under temperature variations ($I_1(T=0) \simeq I_1(T = T_c)$).

In conclusion it must be noted that the photoinduced superlattice of alternating metallic and superconducting phases in a vanadium dioxide film on a substrate, the superlattice observed so far in experiments, constitutes an irreversible structure, with a period $\lambda \sim 1\ \mu$m, left after the light pulse has passed through the system,[19] and probably cannot be interpreted by the theory developed in this paper. Experiments that would detect a thermodynamically nonequilibrium superstructure that disappears after irradiation has ceased have yet to be conducted.

*)E-mail: semenov@quant.univ.simbirsk.su

[1] L. N. Bulaevskiĭ, Usp. Fiz. Nauk **115**, 263 (1975) [Sov. Phys. Usp. **18**, 131 (1975)].
[2] S. N. Artemenko, A. F. Volkov, S. V. Zaĭtsev-Zotov, Usp. Fiz. Nauk **166**, 434 (1996) [Phys. Usp. **39**, 403 (1996)].
[3] L. P. Gor'kov, Usp. Fiz. Nauk **144**, 381 (1984) [Sov. Phys. Usp. **27**, 809 (1984)].
[4] A. A. Bugaev, B. P. Zakharchenya, and F. A. Chudnovskiĭ, *The Metal–Insulator Phase Transition and Its Application* [in Russian], Nauka, Leningrad (1979), pp. 28 and 106.
[5] R. O. Zaĭtsev, E. V. Kuz'min, and S. G. Ovchinnikov, Usp. Fiz. Nauk **148**, 603 (1986) [Sov. Phys. Usp. **29**, 322 (1986)].
[6] A. L. Semenov, Fiz. Tverd. Tela (St. Petersburg) **39**, 925 (1997) [Phys. Solid State **39**, 826 (1997)].
[7] V. I. Emel'yanov, N. L. Levshin, and A. L. Semenov, Vestnik Moskov. Univ. Ser. III Fiz. Astronom. **30**, 52 (1989).
[8] V. I. Emel'yanov, N. L. Levshin, and A. L. Semenov, Fiz. Tverd. Tela (Leningrad) **31**(10), 261 (1989) [Sov. Phys. Solid State **31**, 1803 (1989)].
[9] V. I. Emel'yanov, N. L. Levshin, and A. L. Semenov, Vestnik Moskov. Univ. Ser. III Fiz. Astronom. **31**, 99 (1990).
[10] A. L. Semenov, Fiz. Tverd. Tela (St. Petersburg) **36**, 1974 (1994) [Phys. Solid State **36**, 1079 (1994)].
[11] A. L. Semenov and S. V. Sukhov, Izv. Vyssh. Uchebn. Zaved. Fiz. **6**, 120 (1996).
[12] V. I. Emel'yanov and A. L. Semenov, Fiz. Tverd. Tela (Leningrad) **32**, 3083 (1990) [Sov. Phys. Solid State **32**, 1790 (1990)].
[13] V. I. Emel'yanov, N. L. Levshin, S. Yu. Poroĭkov, and A. L. Semenov, Vestnik Moskov. Univ. Ser. III Fiz. Astronom. **32**, 63 (1991).
[14] K. F. Berggren and V. A. Huberman, Phys. Rev. B **18**, 3369 (1978).
[15] V. V. Kapaev, Yu. V. Kopaev, and S. N. Molotkov, Mikroélektronika **12**, 499 (1983).
[16] V. F. Elesin, V. V. Kapaev, and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **71**, 714 (1976) [Sov. Phys. JETP **44**, 375 (1976)].
[17] A. A. Bugaev, B. P. Zakharchenya, and F. A. Chudnovskiĭ, JETP Lett. **33**, 629 (1981).
[18] A. A. Bugaev, V. V. Gudyalis, B. P. Zakharchenya, and F. A. Chudnovskiĭ, JETP Lett. **34**, 430 (1981).
[19] A. A. Bugaev and A. V. Klochkov, Fiz. Tverd. Tela (Leningrad) **26**, 3487 (1984) [Sov. Phys. Solid State **26**, 2100 (1984)].
[20] V. I. Emel'yanov and I. F. Uvarova, Zh. Éksp. Teor. Fiz. **94**(8), 255 (1988) [Sov. Phys. JETP **67**, 1662 (1988)].
[21] R. F. Mamin, Zh. Éksp. Teor. Fiz. **111**, 1465 (1997) [JETP **84**, 808 (1997)].
[22] Yu. V. Kopaev, V. V. Menyaĭlenko, and S. N. Molotkov, Zh. Éksp. Teor. Fiz. **89**, 1404 (1985) [Sov. Phys. JETP **62**, 813 (1985)].
[23] Yu. V. Kopaev and R. Kh. Timerov, Zh. Eksp. Teor. Fiz. **63**, 290 (1972) [Sov. Phys. JETP **36**, 153 (1973)].

[24] O. Madelung, *Introduction to Solid State Physics*, Springer-Verlag, Berlin (1978), p. 449.
[25] N. N. Bogolyubov and N. N. Bogolyubov Jr., *Introduction to Quantum Statistical Mechanics* [in Russian], Nauka, Moscow (1984), p. 282.
[26] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977), Chap. VI.
[27] N. R. Belashenkov, V. B. Karasev, A. A. Solunin, I. A. Khakhaev, K. Sh. Tsibadze, and F. A. Chudnovskiĭ, Fiz. Tverd. Tela (St. Petersburg) **36**, 2475 (1994) [Phys. Solid State **36**, 1347 (1994)].
[28] V. L. Bonch-Bruevich and S. G. Kalashnikov, *Semiconductor Physics* [in Russian], Nauka, Moscow (1977), p. 255.
[29] S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radiophysics and Optics* [in Russian], Nauka, Moscow (1981), p. 42.
[30] D. N. Klyshko, *Physical Bases of Quantum Electronics* [in Russian], Nauka, Moscow (1986), p. 22.
[31] A. S. Davydov, *Solid-State Theory* [in Russian], Nauka, Moscow (1976), p. 296.
[32] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics: Superradiation, Bistability, and Phase Transitions* [in Russian], Nauka, Moscow (1988), p. 256 [English transl.: *Cooperative Effects in Optics*, IOPP, Bristol (1993)].

Translated by Eugene Yankovsky

Effect of the micromagnetic structure of domain walls on the properties of an isolated domain in a thin magnetic film

Yu. I. Dzhezherya

Institute of Magnetism, National Academy of Sciences, Ukraine
(Submitted 27 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1315–1325 (April 1999)

The effect of the inner structure of domain walls on the time-independent parameters of an isolated stripe domain in a thin ferromagnetic film is studied. The adopted variant of the perturbation theory makes it possible to account, within a unified approach, for the contributions of the magnetostatic and exchange interactions. © *1999 American Institute of Physics.* [S1063-7761(99)01104-X]

The physical theory of magnetic domain walls has been thoroughly studied.[1–5] The current interest can be explained, on the one hand, by the broad promise of using materials containing domains in microelectronics and, on the other, by the development of the nonlinear theory of magnetism.

The classical theory of ferromagnetism is based on the dynamical Landau–Lifshitz equations, which make it possible to extract exhaustive information about the state of the magnetic system. However, if we use models that are close to reality, the equations become very complicated and many simplifying assumptions have to be introduced into the calculations.

For instance, in a model widely used in studies of properties of magnetic domains, the domain walls are interpreted as infinitely thin geometric boundaries with their own surface energy.[1,6–8] This model provides a good description of the state of the system in low magnetizing fields but cannot be used in fields with large amplitudes, when the domain width becomes comparable to the thickness of the domain walls. In such a situation the inner structure of the domain walls has a strong effect on the properties of the magnetic domains.

The present study uses the Landau–Lifshitz equations to investigate an isolated stripe domain localized in a thin ferromagnetic film and the dependence of the domain properties on the inner structure of the domain wall. The limits of the results will also be investigated. The method of regularizing perturbations of the nonlinear equations developed in this paper in general form can be used to study other physical problems.

The characteristics of the system investigated in the present paper are determined by the energy functional with the following structure:

$$E=\int\int dx\,dy\int_0^L dz\,w(\mathbf{m}),$$

$$w(\mathbf{m})=M_0^2\left\{\frac{\alpha}{2}\left(\frac{\partial\mathbf{m}}{\partial\mathbf{x}}\right)^2+\frac{\beta}{2}(1-m_z^2)-h_z m_z-\frac{1}{2}\mathbf{m}\mathbf{h}_m\right\}, \qquad (1)$$

where $\mathbf{m}=\mathbf{M}/M_0$ is the unit magnetization vector, M_0 is the saturation magnetization, α and β are, respectively, the exchange-interaction and uniaxial-anisotropy constants, $h_z=H_z/M_0$ is the reduced magnetic field parallel to the easy-magnetization axis and orthogonal to the plane of the film, $\mathbf{h}_m$ is the intrinsic magnetostatic field of the ferromagnet, and L is the film thickness.

In our domain-structure studies we will assume that the plane of the domain walls limiting the isolated stripe domain is orthogonal to the x axis. It has proved convenient to study the magnetization states in terms of the angular variables θ and φ, which are the polar and azimuthal angles in a system of coordinates whose polar axis is the x axis. Here the relationship between the components of the magnetization vector and the new variables is

$$\mathbf{m}=(\cos\theta,\ \sin\theta\sin\varphi,\ \sin\theta\cos\varphi). \qquad (2)$$

As is known, the magnetostatic field of the sample induced by magnetic inhomogeneities is given by the formula[3]

$$\mathbf{h}_m=\nabla\int d\mathbf{r}'\left(m_i(\mathbf{r}')\frac{\partial}{\partial x_i}\right)\frac{1}{|\mathbf{r}-\mathbf{r}'|}. \qquad (3)$$

Mathematically, magnetic domain walls are represented by the soliton solutions of the Landau–Lifshitz equations. These equations for the angular variables in the time-independent case can be derived by varying the energy functional (1) in the variables θ and φ,

$$\frac{\partial w}{\partial\theta}-\nabla\frac{\partial w}{\partial\nabla\theta}=0,\quad \frac{\partial w}{\partial\varphi}-\nabla\frac{\partial w}{\partial\nabla\varphi}=0, \qquad (4)$$

with the following boundary conditions at the film surface:

$$\frac{\partial\theta(z=0)}{\partial z}=\frac{\partial\theta(z=L)}{\partial z}=\frac{\partial\varphi(z=0)}{\partial z}=\frac{\partial\varphi(z=L)}{\partial z}=0. \qquad (5)$$

When studying the domain structures in film materials, one must allow for the effect of the magnetostatic fields generated by the surface inhomogeneities in the magnetization distribution. The presence of such inhomogeneities distorts the inner structure of the domain walls. However, as shown in Ref. 1, for thin-film materials whose thickness $L\approx\Lambda=\sqrt{\alpha/4\pi}$, these distortions are effectively suppressed by intensive exchange interaction. Hence in the zeroth approximation the structure of the domain walls can be assumed to be

of the Bloch type. The assumption is especially true when one deals with ultrathin magnetic films[9] (the technology needed to fabricate such films is being actively developed). In our notation, this case corresponds to $\theta=\pi/2$. We will assume that deviations from the Bloch configuration can be described by small corrections $\theta_1(\mathbf{r})$, such that $|\theta_1|\ll 1$. The limits of the approximations are examined in the Appendix.

Thus, if we assume that

$$\theta(\mathbf{r})=\frac{\pi}{2}+\theta_1(\mathbf{r}) \tag{6}$$

and that the distribution of magnetization is uniform along the x axis parallel to the plane of the domain walls, we obtain equations that describe the state of the domain structure to within terms of first order in $\theta_1(\mathbf{r})$:

$$-l^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)\varphi+\sin\varphi\cos\varphi+\varepsilon \times(h_z+h_z^m(x,z,\varphi))\sin\varphi=0, \tag{7.1}$$

$$\hat{L}(\varphi)\theta_1-l^2\frac{\partial^2\theta}{\partial z^2}+\varepsilon h_x^m(x,z,\varphi)=0, \tag{7.2}$$

$$\begin{pmatrix} h_x^m \\ h_z^m \end{pmatrix}=-\begin{pmatrix} \dfrac{\partial}{\partial x} \\ \dfrac{\partial}{\partial z} \end{pmatrix}\int_{-\infty}^{\infty}dx'\int_0^L dz'\cos\varphi \times\frac{z-z'}{(x-x')^2+(z-z')^2}, \tag{7.3}$$

where

$$\hat{L}(\varphi)=-l^2\frac{\partial^2}{\partial x^2}+\cos^2\varphi-l^2\left(\frac{\partial\varphi}{\partial x}\right)^2$$

is a linear operator, h_i^m are the components of the magnetostatic field, $l=\sqrt{\alpha/\beta}$ is the domain-wall thickness parameter, and $\varepsilon=1/\beta\ll 1$ is a small parameter. Terms proportional to $\varepsilon\theta_1$ have been dropped from Eq. (7.2).

Obviously, the magnetization inhomogeneities on the surface of the film, generated by the domain walls, lead to the emergence in Eq. (7.1) of a small nonlinear operator $\varepsilon h_z^m(x,z,\varphi)$, which is explicitly coordinate-dependent. The variants of the theory for studying the effect of adiabatic temporal perturbations are discussed in Ref. 10. Here we generalize the results of Ref. 10 to the case of slow spatial modulations.

Since θ_1 is assumed small, in the leading approximation the state of the magnetization field is determined by the boundary value problem

$$-l^2\frac{\partial^2\varphi}{\partial x^2}+\sin\varphi\cos\varphi =l^2\frac{\partial^2\varphi}{\partial z^2}-\varepsilon(h_z+h_z^m(x,z,\varphi))\sin\varphi, \tag{8.1}$$

$$\frac{\partial\varphi(z=0)}{\partial z}=\frac{\partial\varphi(z=L)}{\partial z}=0,\quad \frac{\partial\varphi(x\to\pm\infty)}{\partial x}=0.$$

The variations in φ along the z axis are caused by small perturbations proportional to ε, so that we will assume that these variations are small. The right-hand side of Eq. (8.1) is considered a perturbation.

Ordinary perturbation theory can be used if a perturbation leads to small quantitative corrections. Hence, when studying a nonlinear problem, we must ensure that the perturbation we ignored in the zeroth approximation does not introduce variations of the qualitative nature into the system.

For instance, in a zero magnetic field h_z, one of the ground states of the system involved is a solitary 180° domain wall, while in a finite magnetic field the steady-state solution describing this structure is unstable, and a ground state is, for example, an isolated stripe domain magnetized in the direction opposite to that of the external magnetic field. In this sense the perturbation described by the expression on the right-hand side of Eq. (8.1) is of a singular nature and dramatically transforms the state of the system.

To apply the methods of perturbation theory we must regularize the perturbation.

One variant of regularization can be represented in general form for an arbitrary equations that has soliton solutions.

Let us consider a boundary value problem for the nonlinear equation

$$-l^2\frac{\partial^2\varphi}{\partial x^2}+\hat{L}_0(\varphi)=l^2\frac{\partial^2\varphi}{\partial z^2}-\varepsilon h(x,z,\varphi), \tag{9}$$

where $\hat{L}_0(\varphi)$ and $h(x,z,\varphi)$ are nonlinear operators, $\varepsilon\ll 1$ is a small parameter whose modulations along the z axis are due to a perturbation and are assumed small.

For the sake of definiteness we assume that the boundary conditions are

$$\frac{\partial\varphi(x\to\pm\infty)}{\partial x}=0, \tag{10.1}$$

$$\frac{\partial\varphi(z=0,L)}{\partial z}=0. \tag{10.2}$$

We also assume that the perturbation is of a singular nature and cannot be take into account by standard perturbation-theory methods. We will attempt to single out the specific feature of this perturbation by specifying a simpler operator $\varepsilon L_1(H,\varphi)$, where H is an undefined parameter, which generally depends on the variable z.

We introduce this operator into Eq. (9), which becomes

$$-l^2\frac{\partial^2\varphi}{\partial x^2}+\hat{L}_0(\varphi)+\varepsilon\hat{L}_1(H,\varphi) =l^2\frac{\partial^2\varphi}{\partial z^2}-\varepsilon(h(x,z,\varphi)-\hat{L}_1(H,\varphi)). \tag{11}$$

We performed this transformation in order to select an effective operator $L_1(H,\varphi)$ that affects the structure of the solution in the same as the initial perturbation $h(x,z,\varphi)$.

Thus, assuming that the right-hand side of Eq. (11) is the regularized perturbation, we write the solution of (11) as follows:

$$\varphi(x)=\varphi_0(x-x_0,H)+\varepsilon\varphi_1(x-x_0,H)+\cdots, \tag{12}$$

where $\varphi_0(x-x_0,H)$ is the two-parameter solution of the boundary value problem

$$-l^2\frac{\partial^2\varphi_0}{\partial x^2}+L_0(\varphi_0)+\varepsilon L_1(H,\varphi_0)=0, \tag{13}$$

$$\frac{\partial\varphi_0(x\to\pm\infty)}{\partial x}=0.$$

Since Eq. (13) does not depend explicitly on the variable x, its solution can be obtained in general form at least as integrals. The solution is translation invariant.

The value of the correction is determined by the inhomogeneous boundary value problem

$$G_0(\varphi_0,H)\varphi_1=f_1(X,x_0,z,H),\quad \varphi_1(x=\pm\infty,H)=0, \tag{14}$$

$$f_1(X,x_0,z,H_0)=(L_1(\varphi_0,H_0)-h(X+x_0,z,\varphi_0))+\alpha\frac{\partial^2\varphi_0}{\partial z^2},$$

$$\frac{\partial^2\varphi_0}{\partial z^2}=\frac{\partial^2\varphi_0}{\partial x^2}\left(\frac{\partial x_0}{\partial z}\right)^2-\frac{\partial\varphi_0}{\partial x}\frac{\partial^2x_0}{\partial z^2}-2\frac{\partial^2\varphi_0}{\partial x\,\partial H}\frac{\partial x_0}{\partial z}\frac{\partial H}{\partial z}+\frac{\partial^2\varphi_0}{\partial H^2}\left(\frac{\partial H}{\partial z}\right)^2+\frac{\partial\varphi_0}{\partial H}\frac{\partial^2H}{\partial z^2},$$

$$G_0(\varphi_0,H)=-l^2\frac{\partial^2}{\partial X^2}+\frac{\partial L_0(\varphi_0,H)}{\partial\varphi_0}+\varepsilon\frac{\partial L_1(\varphi_0,H)}{\partial\varphi_0}, \tag{15}$$

where $X=x-x_0$.

For the right-hand side not to contain secular terms, we must correctly determine the parameters x_0 and H. This is achieved by ensuring that the inhomogeneous equation (14) meets the solvability conditions, and this requires knowing the solutions of the corresponding homogeneous equation. These conditions can easily be found by differentiating Eq. (13) with respect to the parameters H and x_0. Here we have

$$G_0(\varphi_0,H)\psi_1(X)=0,$$

$$G_0(\varphi_0,H)\psi_2(X)=-\varepsilon H\frac{\partial}{\partial H}L_1(\varphi_0,H), \tag{16}$$

where

$$\psi_1(X)=\frac{\partial\varphi_0}{\partial X},\quad \psi_2(X)=H\frac{\partial\varphi_0}{\partial H}.$$

Thus, $\psi_1(x)$ has a zero eigenvalue, is a solution of the homogeneous equation corresponding to (14), and is localized near the solution. Clearly, in the leading approximation in ε these features are also inherent in $\psi_2(x)$.

The solvability conditions for Eq. (14) have the form

$$\int_{-\infty}^{\infty}dx\begin{pmatrix}\psi_1(X)\\ \psi_2(X)\end{pmatrix}f_1(X,x_0,H_0)=0. \tag{17}$$

They determine a system of differential equations that together with the boundary conditions

$$\frac{\partial x_0(z=0,L)}{\partial z}=\frac{\partial H(z=0,L)}{\partial z}=0 \tag{18}$$

make it possible to find the effective parameters.

Since it is known that $\psi_1(X)$ is a solution of the homogeneous equation in (16), we may formally assume that φ_1 has been determined.[11]

The application of this approach simplifies substantially if we wish to find the solution of the equation in the one-dimensional case. Here x_0 and H are constants whose values are determined by the conditions (17). In this case we an easily establish the nature of the perturbation.

If

$$\int_{-\infty}^{\infty}dx\frac{\partial\varphi_0}{\partial x}h(x,\varphi_0)\neq0$$

for all values of the parameter x_0, where φ_0 is the solution of the unperturbed problem

$$-l^2\frac{\partial^2\varphi_0}{\partial x^2}+L_0(\varphi_0)=0,\quad \frac{\partial\varphi_0(x\to\pm\infty)}{\partial x}=0,$$

the solution cannot be approximated by a one-parameter solution and the perturbation is of a singular nature, which requires allowing for its singularity.

The success of this method depends on how we select the effective operator $L_1(\varphi,H)$. Since there is always a certain ambiguity in selecting this operator, a simple form of the operator is preferable.

Following the theory, we select the effective operator in the form $L_1(\varphi,H)=\pm H\sin\varphi$.

Then the leading approximation to the solution of Eqs. (7.1), (8.1) and (8.2) is determined by the simpler boundary value problem

$$-l^2\frac{\partial^2\varphi_0}{\partial x^2}+\sin\varphi_0\cos\varphi_0\pm\varepsilon H\sin\varphi_0=0,$$

$$\frac{\partial\varphi_0(\pm\infty,H)}{\partial X}=0. \tag{19}$$

When H is positive, depending on the sign in front of it the solution of (19) is

$$\varphi_0(X,H)=\begin{cases}\pi+2\arctan\left\{\sqrt{\dfrac{\varepsilon H}{1+\varepsilon H}}\sinh\left(\sqrt{1+\varepsilon H}\dfrac{X}{l}\right)\right\}, & +H,\\ \pi-2\arctan\left\{\sqrt{\dfrac{\varepsilon H}{1-\varepsilon H}}\cosh\left(\sqrt{1-\varepsilon H}\dfrac{X}{l}\right)\right\}, & -H.\end{cases} \tag{20}$$

From the physical viewpoint, the solutions in (20) correspond to bound states of two unipolar and heteropolar Bloch walls. The polarity is determined by the sense of rotation of the magnetization vector in the plane of the domain walls. The parameter H in this case is an effective superposition of the external magnetic field and the intrinsic magne-

tostatic field. Clearly,[12] the value of H is linked to the width parameter d of the stripe domain through the relationship

$$H(z)=4\beta \exp(-d(z)/l). \tag{21}$$

The corrections to the leading approximation are determined by the boundary value problem (14), where

$$G_0(\varphi_0,H)=-l^2\frac{\partial^2}{\partial X^2}+\cos 2\varphi_0\pm\varepsilon H\cos\varphi_0, \tag{22}$$

$$f_1(X,z,x_0,H)=\sin\varphi_0(\pm H-h_z-h_z^m \times(X+x_0,z,\varphi_0))+\alpha\frac{\partial^2\varphi_0}{\partial z^2}.$$

In what follows, a plus sign denotes an isolated stripe domain limited by unipolar domain walls and a minus sign denotes an isolated stripe domain limited by heteropolar domain walls.

In the leading approximation, the solutions ψ_1 and ψ_2 of the homogeneous equation are

$$\psi_1(X)=\frac{\partial\varphi_0}{\partial X},\quad \psi_2(X)=\sin\varphi_0+O(\varepsilon). \tag{23}$$

In our case, instead of deriving an equation for the parameter H, it is convenient to use (17) and derive an equation for the width parameter of an isolated stripe domain, $d(z)$, via the unambiguous relationship (21).

To obtain the equations for the parameters x_0 and d from the conditions (17), it is advisable to use the properties of the function $\varphi_0(X,H)$, which make it possible to evaluate the improper integrals via asymptotic methods to high accuracy.

Clearly, the following conditions hold for an arbitrary smooth function $F(x/L)$:

$$\int_{-\infty}^{\infty}dx\frac{\partial\varphi_0}{\partial x}\sin\varphi_0 F\left(\frac{x}{L}\right) \approx 2\left(F\left(\frac{-d/2+x_0}{L}\right)-F\left(\frac{d/2+x_0}{L}\right)\right)+O\left(\frac{l}{L}\right),$$

$$\int_{-\infty}^{\infty}\frac{dx}{l}\sin^2\varphi_0 F\left(\frac{x}{L}\right) \approx 2\left(F\left(\frac{-d/2+x_0}{L}\right)+F\left(\frac{d/2+x_0}{L}\right)\right)+O\left(\frac{l}{L}\right),$$

$$\int_{-\infty}^{\infty}dx(1-\cos\varphi_0)F\left(\frac{x}{L}\right) \approx 2\int_{-d/2+x_0}^{d/2+x_0}dx F\left(\frac{x}{L}\right)+O\left(\frac{l}{L}\right). \tag{24}$$

Using these properties, we can write the components of the magnetostatic field in the form

$$h_i^m(x,z)=2\frac{\partial}{\partial x_i}\int_0^L dz'\left(\arctan\frac{d(z')/2+x-x_0(z')}{z-z'} +\arctan\frac{d(z')/2-x+x_0(z')}{z-z'}\right),\quad i=x,z. \tag{25}$$

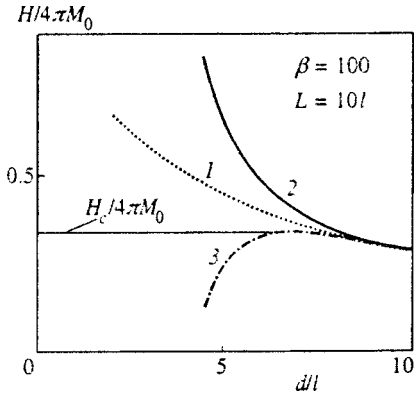

FIG. 1.

Thus, on the basis of Eqs. (21) and (23)–(25) and the conditions (17) we obtain

$$l\frac{\partial^2 x_0}{\partial z^2}=\frac{\varepsilon}{2}\left(h_z^m\left(x_0-\frac{d}{2},z\right)-h_z^m\left(x_0+\frac{d}{2},z\right)\right), \tag{26.1}$$

$$l\frac{\partial^2 d}{\partial z^2}=\varepsilon\left(\pm 8\beta\exp\left(-\frac{d}{l}\right)-2h_z-h_z^m\left(x_0-\frac{d}{2},z\right) -h_z^m\left(x_0+\frac{d}{2},z\right)\right). \tag{26.2}$$

Clearly, $x_0=\text{const}$ is an exact solution of Eq. (26.1) corresponding to the boundary conditions (18). Equation (26.2) is a nonlinear integrodifferential equation and cannot be solved exactly. However, since its right-hand side contains a small parameter and is actually a perturbation, the solution (in accordance with the boundary conditions $\partial d(0)/\partial z=\partial d(L)/\partial z=0$ can be written $d(z)=d_0+d_1(z)+\cdots$, where d_0 is a constant whose value has yet to be found.

From Eq. (26.2) it follows that the equation for the corrections in the approximation linear in ε is

$$l\frac{\partial^2 d}{\partial z^2}=\varepsilon\left\{-2h_z\pm 8\beta\exp\left(-\frac{d_0}{l}\right)-8 \times\left(\arctan\frac{d_0}{L-z}+\arctan\frac{d_0}{z}\right)\right\}. \tag{27}$$

By integrating Eq. (27) over the interval $[0<z<L]$ with allowance for the boundary conditions we obtain a relationship the links the width of an isolated stripe domain to the strength of the external magnetic field:

$$h_z=4\left\{\pm\beta\exp(-d_0/l)+2\arctan\frac{L}{d_0} -\frac{d_0}{L}\ln\left(1+\left(\frac{L}{d_0}\right)^2\right)\right\}. \tag{28}$$

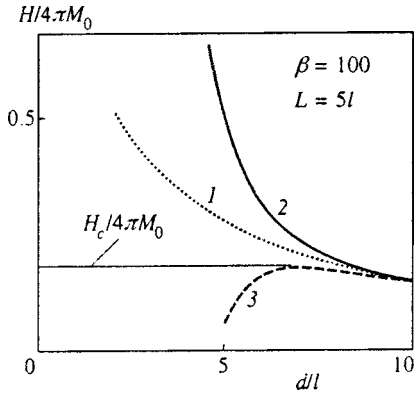

FIG. 2.

The first term on the right-hand side corresponds to the exchange interaction due to allowance for the inner structure of the stripe domain. It is dominant over a narrow stripe domain and is negligible in weak magnetic fields, when (28) asymptotically tends to the limit obtained in the model of geometric domain boundaries.[1] Figures 1 and 2 depict, for films of different thicknesses, the dependence of the width of a stripe domain in the model of geometric domain walls (curve *1*) and of a domain limited by unipolar and heteropolar domain walls (curves *2* and *3*, respectively).

The results obtained in this paper make it possible to establish the value of the critical magnetic field above which there can be no bound state of heteropolar Bloch walls. Obviously, the critical field is defined as the extremum of the function (28) for heteropolar domain walls, whose plots are depicted in Figs. 1 and 2. The dependence of the critical field on the film thickness for materials with different values of anisotropy is depicted in Fig. 3.

The solution of Eq. (27) can easily be expressed in terms of elementary functions. The width of the domain is the greatest at the center of the magnetic film. However, according to estimates for materials with an anisotropy $\beta=100$ and a thickness $L\approx 10l$, the correction d_1 is very small and does exceed $0.1l$. Calculations show that in examining an isolated stripe domain one can ignore the distortions in its structure over the thickness for a broad class of films.

Thus, the proposed method of regularizing perturbations has proved extremely useful in studies of an applied physical problem. The method has made it possible to transcend the geometric domain-wall model and investigate the state of a magnetic domain over a range of field strengths in which the inner structure of the domain walls determines (together with the magnetostatic interaction) the property of the system.

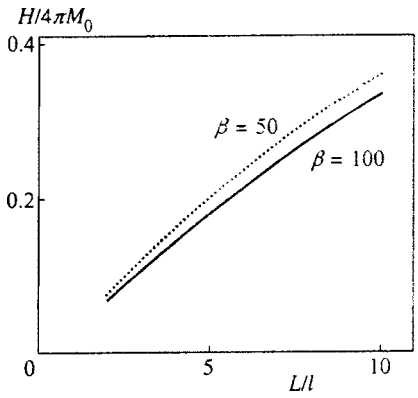

FIG. 3.

I am grateful to V. G. Bar'yakhtar and Yu. I. Gorobets for useful discussions of the material contained in this paper.

APPENDIX

In this appendix we discuss the use of the Bloch-wall approximation. The obvious criterion of validity of the adopted assumptions is that $|\theta_1|\ll 1$, where of θ_1 is the solution of the boundary value problem obtained by substituting $\varphi(x)\approx\varphi_0(x,H)$ into (7.2) and performing the necessary transformations:

$$G_0(\varphi_0,H)\,\theta_1 - l^2\frac{\partial^2\theta_1}{\partial z^2} = -\varepsilon h_x^m(x,z), \tag{A1}$$

$$h_x^m(x,z) = -\ln\frac{(L-z)^2+(x-d/2)^2z^2+(x+d/2)^2}{(L-z)^2+(x+d/2)^2z^2+(x-d/2)^2},$$

$$\frac{\partial\theta_1(z=0,L)}{\partial z}=0.$$

Here we have dropped terms proportional to $\varepsilon\theta_1$ and used the condition $l^2(\partial\varphi_0/\partial x)^2=\sin^2\varphi_0+O(\varepsilon)$.

Since the eigenfunctions (23) of the discrete low-lying levels of the operator $G_0(\varphi_0,H)$ are known, we can expand $\theta_1(x,z)$ in these functions:

$$\theta_1(x,z)=C_1(z)l\frac{\partial\varphi_0}{\partial x}+C_2(z)\,\sin\,\varphi_0. \tag{A2}$$

Plugging (A2) into Eq. (A1) and finding the scalar products of the result and the corresponding eigenfunctions of the discrete spectrum, we obtain equations for the expansion coefficients $C_i(z)$, which for an isolated stripe domain limited by unipolar domain walls obey the relationship

$$\frac{\partial^2 C_2}{\partial z^2}=\frac{1}{2\Lambda^2}\left\{\ln\frac{z}{L-z}-\frac{1}{2}\ln\frac{d_0^2+z^2}{d_0^2+(L-z)^2}\right\} \tag{A3}$$

and for an isolated stripe domain limited by heteropolar domain walls, the relationship

$$\frac{\partial^2 C_1}{\partial z^2}=\frac{1}{2\Lambda^2}\left\{\ln\frac{z}{L-z}-\frac{1}{2}\ln\frac{d_0^2+z^2}{d_0^2+(L-z)^2}\right\}, \tag{A4}$$

where $\Lambda=\sqrt{\alpha/4\pi}$ is the characteristic magnetic length.

The first term in the braces describes the effect of the intrinsic magnetostatic field on the magnetization distribution in the domain walls limiting the isolated stripe domain. The second term describes the influence of an adjacent domain wall on the magnetization state and moderates somewhat the magnetization twisting effects in the domain walls (analysis shows that the contribution of this term is insignificant). Hence in estimating the perturbation we ignore the second term, which only strengthens the criterion.

The value of the angle $\theta_1(x,z)$ specified by Eqs. (A3) and (A4) and the relationship (A2) has the form

$$\theta_1(x,z)\approx\begin{cases}\gamma(z)\ \sin\ \varphi_0(x),\\ \gamma(z)l\dfrac{\partial\varphi_0(x)}{\partial x},\end{cases} \tag{A5}$$

with the upper value corresponding to a bound state of the unipolar domain walls limiting the isolated stripe domain and the lower value, a bound states of the heteropolar domain walls limiting the isolated stripe domain. The factor $\gamma(z)$ is given by the formula

$$\gamma(z)=\frac{1}{4}\left(\frac{L}{\Lambda}\right)^2\left(\left(\frac{z}{L}\right)^2\ln\frac{z}{L}-\left(1-\frac{z}{L}\right)^2\ln\left(1-\frac{z}{L}\right)+\frac{1}{4}\left(1-\frac{2z}{L}\right)\right).$$

The value of θ_1 is at its maximum at the surface of the film:

$$|\theta_{1\ \max}|=\frac{1}{16}\left(\frac{L}{\Lambda}\right)^2.$$

Thus, the configuration of the domain walls in thin films with parameters satisfying the condition $(L/\Lambda)^2\ll 16$ may be considered to be of the Bloch type.

[1] A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials*, Applied Solid State Science Series, Supplement I, Academic Press, New York (1979).
[2] A. M. Kosevich, B. A. Ivanov, and A. S. Kovalev, *Nonlinear Magnetization Waves. Dynamic and Topological Solitons* [in Russian], Naukova Dumka, Kiev (1983).
[3] V. G. Bar'yakhtar and Yu. I. Gorobets, *Magnetic Bubbles and Their Lattices* [in Russian], Naukova Dumka, Kiev (1988).
[4] A. H. Bobeck and E. Della Torre, *Magnetic Bubbles*, North-Holland, Amsterdam (1975).
[5] A. Hubert, *Theorie der Domanenwande in geordneten Medien*, Springer, Berlin (1974).
[6] V. G. Bar'yakhtar, Yu. I. Gorobets, and S. I. Denisov, Ukr. Fiz. Zh. **28**, 436 (1983).
[7] F. L. Vaĭsman, Yu. I. Gorobets, and S. I. Denisov, Ukr. Fiz. Zh. **31**, 1234 (1986).
[8] A. P. Babichev, Yu. I. Gorobets, and S. I. Denisov, Ukr. Fiz. Zh. **30**, 86 (1985).
[9] A. L. Sukstanskiĭ and V. V. Tarasenko, Zh. Éksp. Teor. Fiz. **112**, 1476 (1997) [JETP **85**, 804 (1997)].
[10] *Solitons in Action*, K. Lonngren and A. Scott (eds.), Academic Press, New York (1978).
[11] E. Kamke, *Handbook of Ordinary Differential Equations*, Nauka, Moscow (1976) [Russian translation of *Differentialgleichungen, Lösungsmethoden und Lösungen*, Vol. I, Chelsea, New York (1948)].
[12] Yu. I. Dzhezherya, Fiz. Tverd. Tela (St. Petersburg) **35**, 2270 (1993) [Phys. Solid State **35**, 1366 (1993)].

Translated by Eugene Yankovsky

EPR observations of anomalous triplet states in $Ba_{1-x}K_xBiO_3$ and $BaPb_yBi_{1-y}O_3$

A. Yu. Yakubovskiĭ and S. V. Gudenko

Kurchatov Institute, 123182 Moscow, Russia

N. V. Anshukova and A. I. Golovashkin

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

L. I. Ivanova and A. P. Rusakov

Moscow Institute of Steel and Alloys, 117049 Moscow, Russia

(Submitted 3 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1326–1336 (April 1999)

Electron paramagnetic resonance (EPR) spectra of samples of the systems $Ba_{1-x}K_xBiO_3$ and $BaPb_yBi_{1-y}O_3$ are investigated over wide ranges of composition and temperature. Two main lines in the EPR spectrum with factors $g_1 \approx 2.1$ and $g_2 \approx 4.2$ are found for all compositions. It is shown that the observed EPR line with $g_2 \approx 4.2$ is due to oxygen ions. This probably indicates the presence of oxygen ions with different effective charges, i.e., the existence of charge density waves in the oxygen-ion sublattice in addition to charge density waves in the bismuth sublattice. © *1999 American Institute of Physics.* [S1063-7761(99)01204-4]

1. INTRODUCTION

Compounds based on bismuth—$Ba_{1-x}K_xBiO_3$ (BKBO) and $BaPb_yBi_{1-y}O_3$ (BPBO)—with perovskite structure are similar in their crystalline structure and in a number of physical properties to cuprate high-temperature superconductors (HTSC). The relatively high critical temperatures T_c of these compounds, despite the lack of copper ions and the magnetic moments associated with them, make these systems a very interesting object for checking various hypotheses about the nature of HTSC's. The anomalous behavior of some of the physical properties of BKBO and BPBO systems, and the nature of the metal–insulator concentration transition and superconductivity in these compounds, have yet to be thoroughly explained.

A number of unusual properties of BKBO and BPBO, and also of cuprate HTSC's, have been linked to structural, charge, or magnetic ordering in the Bi or Cu sublattice. However, the observed anomalous softening of longitudinal optical phonon frequencies in the [100] direction for the metallic phases in BKBO (Ref. 1) and in the cuprate HTSC's $La_{2-x}Sr_xCuO_4$ (LSCO) and $YBa_2Cu_3O_{7-\delta}$ (YBCO) (Ref. 2) cannot be explained by charge ordering in the bismuth-ion sublattice or by spin ordering in the copper-ion sublattice. Ordering of this kind should lead to anomalies in the phonon spectrum in the [110] directions, which are not observed in experiment.[1,2]

In all of these compounds an anomalous (negative) thermal expansion is observed at low temperatures,[3,4] as well as an anomalous temperature dependence of the upper critical magnetic field H_{c2} (Refs. 5 and 6) along with other anomalies. In BKBO, $Nd_{2-x}Ce_xCuO_4$, YBCO, the bismuth cuprate HTSC's $Bi_2Sr_2CuO_{6+\delta}$, and $Bi_2Sr_2CaCu_2O_{8+\delta}$, the cross section of the Fermi surface is almost a square[7–11] with flat segments parallel to the [100] directions. The singling out of these directions can also not be explained simply by ordering in the copper-ion sublattice for cuprate HTSC's or in the bismuth-ion sublattice in BKBO since, as already noted, the symmetric directions for all orderings are the directions [110].[12]

Taking these and other experimental facts into account, we conjectured[13,14] that in HTSC systems, besides the orderings with symmetry in the [110] directions, there exists a charge ordering in the oxygen-ion sublattice, i.e., a charge density wave in this sublattice with symmetry directions [100]. In particular, for BKBO (Ref. 13), besides the charge density wave in the [110] direction in the bismuth-ion sublattice, it is probably also necessary to allow for a charge density wave in the oxygen-ion sublattice in the [100] direction. In the approximation of the tight-binding method this means that oxygen ions with different effective charges should exist in the system, whose ordering is important for the formation of the electron band structure and the Fermi surface.

A preliminary series of experiments carried out by us, and in particular measurements of EPR spectra,[15] show that in BKBO ordering of oxygen ions with different effective charges possibly exists. For a more detailed elucidation of the nature of the observed EPR signals and the role of oxygen, in the present work we carried out integrated studies of BKBO and BPBO systems. The measurements were carried out on samples of varied composition, where the concentrations of all the cations (Ba, Bi, K, Pb) and oxygen varied over a wide range. Special attention was given to the quality and certification of the samples.

2. SAMPLES

Samples of the system $Ba_{1-x}K_xBiO_3$ ($0 \leqslant x \leqslant 0.50$) were prepared by nitrate technology.[16] Stoichiometric quantities of

KNO_3, $Ba(NO_3)_2$, and Bi_2O_3 powders were mixed and heated in a nitrogen atmosphere at 965 K for 1 h, and then at 988 K for 30 min. The mixture was cooled to 720 K, the nitrogen atmosphere was replaced by an oxygen atmosphere, and the mixture was kept in it for 30 min. The temperature was then dropped to 420 K and the mixture was taken out of the oven. A similar cycle of ''grinding–synthesis–annealing'' was carried out five times for each composition. Afterwards the samples were pressed into pellets which were held at 988 K for 1 h in a nitrogen atmosphere, and then at 720 K for 30 min in an oxygen atmosphere, then slowly (10 h) cooled to 420 K and taken out of the oven.

X-ray diffraction revealed a single-phase character of the samples with $x \leqslant 0.50$ and a lack of splittings of reflections in the diffraction patterns. The pseudocubic lattice parameter ''a'' had a dependence on the composition x of the form $a = 4.3548 - 0.1743x$ (in Å) in agreement with Ref. 16. To characterize the samples, we measured the temperature dependence of their electrical resistance $R(T)$ and magnetic susceptibility $\chi(T)$. For $x < 0.3$ the samples had a semiconductor dependence of $R(T)$, in the range $0.3 < x \leqslant 0.50$ they manifested metallic properties: in this case the maximum superconducting transition temperature $T_c \approx 30$ K was obtained for the $Ba_{0.6}K_{0.4}BiO_3$ sample. An indication of the high quality of the samples is provided by the large fraction of the Meissner phase (more than 50%) measured in a magnetic field of 4 Oe.

Samples of the system $BaPb_yBi_{1-y}O_3$ were also prepared by nitrate technology, but in substantially different thermal regimes. A stoichiometric mixture of $Pb(NO_3)_2$, $Ba(NO_3)_2$, and Bi_2O_3 powders was heated in an oxygen atmosphere to 950 K and kept at that temperature for 1 h. The temperature was then raised to 990 K, and the mixture was held there for 4 h. The powders were then ground and pressed into pellets which were kept for 4 h in an oxygen atmosphere at 1100 K. The pellets were ground up again, pressed, and kept in an oxygen atmosphere for 4 h at 1200 K. After this, the samples were cooled to 1073 K and taken out of the oven. This procedure was repeated two more times, but at a synthesis temperature of 1300 K, the first time for 4 h, and the second time for 50 h. The last anneal was carried out at 1073 K for 12 h, after which the samples were cooled outside the oven in an oxygen atmosphere. Note that samples with $y \geqslant 0.5$ were prepared in the indicated regime; for samples with smaller values of y, the maximum synthesis temperature was reduced to 1100 K as y was reduced to zero. All remaining conditions of synthesis were identical for all compositions.

As shown in Refs. 17 and 18, annealing of samples of $BaPb_yBi_{1-y}O_{3-\delta}$ in an oxygen flux at 1073 K for 12 h ensures complete oxygen stoichiometry ($\delta = 0.00 \pm 0.01$).

According to x-ray diffraction results, all prepared samples were single-phase. The samples with $y < 0.5$ had a monoclinc structure (for $BaBiO_3$ we have $a = 0.6187$ nm, $b = 0.6138$ nm, $c = 0.8670$ nm, $\beta = 90.165°$). For the samples with $y > 0.5$ an orthorhombic structure was observed (for $BaPb_{0.75}Bi_{0.25}O_3$ we have $a = 0.6079$ nm, $b = 0.6061$ nm, $c = 0.8554$ nm, $\beta = 90°$). To characterize the samples we measured the temperature dependence of their electrical resistance $R(T)$ and magnetic susceptibility $\chi(T)$. For $y < 0.65$ the samples had a semiconductor dependence of $R(T)$, and for $y > 0.65$ they exhibited metallic properties, where the maximum superconducting transition temperature $T_c \approx 12.5$ K obtained for the composition $BaPb_{0.75}Bi_{0.25}O_3$.

Analysis of the samples using a LAMMA-1000 laser microprobe mass-spectrometer with a sensitivity threshold of 10^{17} cm^{-3} did not reveal even a trace of copper or any magnetic impurities.

3. RESULTS

EPR spectra were measured with a Bruker ESP-300 spectrometer at 9.45 GHz in the temperature range 3–300 K. The first derivative of the absorption signal in the magnetic field was measured. We used an ESR-900 flow-through helium cryostat with an Oxford Instruments ITC-4 temperature regulator, for which the measurement error of the temperature in the range $T < 30$ K was at most ± 0.2 K.

We examined samples with the following compositions: the insulator $BaBiO_3$, the ''parent'' compound of both systems, BKBO and BPBO; samples of the $Ba_{1-x}K_xBiO_3$ system with $x = 0.13$—an insulator, $x = 0.30$—a composition near the insulator–metal transition, $x = 0.35$—lightly doped ($T_c \sim 20$ K), $x = 0.40$—optimally doped ($T_c = 30$ K), $x = 0.45$ and $x = 0.50$—heavily doped superconductors (for $x = 0.50$ $T_c \leqslant 15$ K); samples of the $BaPb_yBi_{1-y}O_3$ system with $y = 0.20$ and $y = 0.50$—insulators, $y = 0.65$—a composition near the edge of the insulator–metal transition, $y = 0.75$ ($T_c = 12.5$ K), $y = 0.90$ ($T_c = 8$ K) and $y = 1.00$—metals. In addition, we prepared samples of the systems BKBO and BPBO, specially annealed in argon to create oxygen vacancies.

To eliminate the ''sample degradation'' effect on the measurement results noted in Refs. 19 and 20, all of the basic EPR measurements were made on freshly prepared materials, which were ground into powder immediately after final heat treatment and poured together with pure mineral oil directly into the EPR cells. Subsequent studies were carried out in these cells. Such ''hermetization'' of the samples made it possible not only to eliminate the dependence of the results on the time elapsed since sample preparation, but also to eliminate the possibility of partial orientation of the powder in the magnetic field. In the control samples, which were intentionally left in powder form and not mixed with mineral oil, we observed considerable degradation of the EPR signals, especially the EPR line in the ''half-field'' (see below), only a few days after grinding the pellets. This effect indicates a possible reason for the irreproducibility of results noted in a number of published works. As our preliminary measurements have shown,[13,15] one of the principal reasons for sample degradation is loss of oxygen. Note also that the EPR signal of high-purity mineral oil was very low in comparison with the sample signal, although a weak narrow line with $g \approx 2.0$ was also detected.

Above all, note that the presence of two prominent lines in the EPR sample was typical of all samples. One of the lines was observed in a field $B_0 \approx 0.34$ T ($g_1 \approx 2.1$), the second was observed in the ''half-field'' $B_0 \approx 0.17$ T (g_2

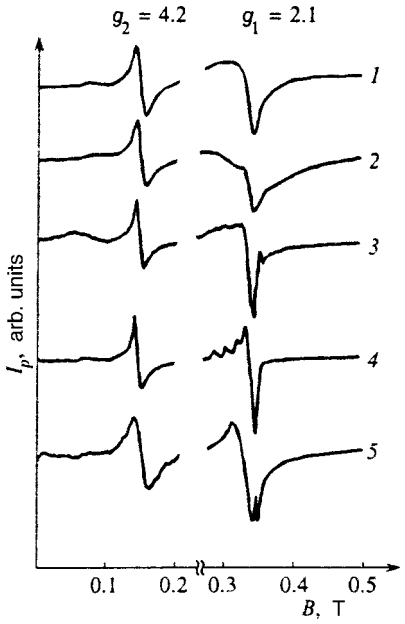

FIG. 1. Examples of EPR lines with $g_1 \approx 2.1$ and $g_2 \approx 4.2$ for samples of the $BaPb_yBi_{1-y}O_3$ and $Ba_{1-x}K_xBiO_3$ systems in a magnetic field B up to 0.5 T: *1*—$y=0$, $T=3.5$ K, *2*—$y=0.5$, $T=3.5$ K, *3*—$y=0.75$, $T=13.5$ K, *4*— $y=1$, $T=3.5$ K, *5*—$x=0.13$, $T=3.5$ K (T is the measurement temperature of the samples).

≈ 4.2). By way of example, Fig. 1 displays spectra for four samples of the $BaPb_yBi_{1-y}O_3$ system ($y=0,0.5,0.75,1.0$) and the $Ba_{0.87}K_{0.13}BiO_3$ sample. As a rule, the total intensity of the ''high-field'' line (g_1) is significan tly greater than that of the ''low-field'' line (g_2), but in some samples their intensities are comparable. It is possible that this depends on the conditions of the anneal or on subtle features of the effect itself. Figure 1 shows the intensities of both lines for all samples, expressed on the same scale and referenced approximately to a common origin.

The intensity (amplitude) of the g_2 line of the samples depended on the temperature T; however, the width of this line and its position did not. For the g_1 line we observed a definite dependence of the position and width on T in the high temperature region ($T>100$ K). Its shape was also observed to evolve with increasing lead content in the BPBO samples.

In the present work we discuss the behavior of only the low-field line with $g_2 \approx 4.2$, which is of greatest interest. The EPR absorption line at half the field strength is well known[21] to indicate the presence of triplet states in the system, i.e., of pairs of states with triplet ordering of the magnetic moments. The most probable source of this EPR signal in the investigated samples is triplet excited states of localized hole pairs with spin $S=1$ (Ref. 15). It follows from the estimated intensity of the lines with $g_2 \approx 4.2$ that the number of such pairs in the samples amounts to $10^{-3}-10^{-2}$ of the number of Bi(Pb) ions. This estimate is consistent with magnetic susceptibility measurements.[22]

The intensities of the two lines fall off rapidly with increasing temperature. Figure 2 shows the temperature dependence of the inverse intensity I_p^{-1} of the EPR lines with $g_2 \approx 4.2$ for several samples of the BPBO and BKBO systems (here I_p is the peak-to-peak amplitude of the EPR line). It can be seen that the results for all samples at $T<30-40$ K are well described by the dependence $I_p^{-1} \propto \chi_{EPR}^{-1}$, where $\chi_{EPR}^{-1}=C/(T+\Theta)$, with a common value of Θ for samples of different composition in Ba, K, Bi, and Pb, but the same stoichiometric oxygen content. Thus, the nature and intensity of the interaction between localized pairs, defined by the parameter Θ, depend weakly on the potassium and lead doping levels of the systems considered. The mean value of the parameter $\Theta=4\pm2$ K.

Since there are several possibilities in BPBO and BKBO systems for the formation of localized triplet pairs, we carried out a series of measurements on the effect of doping and heat treatment on the observed EPR signals. We observed an EPR signal in the half-field in the parent material $BaBiO_3$; consequently, we know that this is signal is not due to potassium or lead ions. On the other hand, it follows from the bismuth substitution experiments carried out in this study that these signals are not due to bismuth ions either. For partial replacement of barium by potassium (up to 50%) the EPR signal varied only slightly. Taking into account here that the EPR signals varied strongly (as will be made clear below) only when the oxygen content was varied, it is natural to assume that these signals are due to oxygen ions.

To clarify the role of oxgyen in the formation of the observed signals, we carried out an additional series of experiments. High-quality BKBO and BPBO samples were annealed in argon at $T=1070$ K for 1 h and then quenched. The measurements showed that the amplitude of the signal with $g_2 \approx 4.2$ decreased roughly tenfold with some difference in the magnitude of decrease for different compositions. Next, the BPBO samples were annealed again, this time in an oxygen atmosphere at $T=1070$ K for 1 h, and then quenched. In

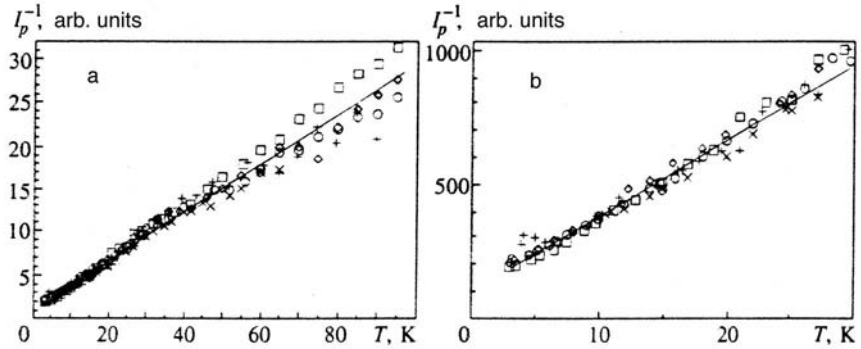

FIG. 2. Temperature dependence of the inverse intensity I_p^{-1} of the EPR line with $g_2 \approx 4.2$ for the $BaPb_yBi_{1-y}O_3$ systems and $Ba_{1-x}K_xBiO_3$ systems: $y=1$ (○), $y=0.75$ (×), $y=0.5$ (□), $y=0.2$(◇), $x=0.4$ (+) (the BKBO sample annealed in argon, $T_c<4$ K). The solid line shows the dependence $I_p^{-1}(T)=(T+\Theta)/C$. a) The dependence $I_p^{-1}(T)$ up to $T=100$ K; b) the dependence $I_p^{-1}(T)$ in the lower temperature region ($T<30$ K).

this case the amplitude of the EPR signals increased severalfold. The incomplete recovery of the EPR signal to its original value is due to insufficient duration of the oxygen anneal. Complete recovery of the oxygen stoichiometry would require a 12-hour anneal.[17,18] However, such a lengthy high-temperature anneal would lead to other changes in the samples associated with their preparation. A corresponding anneal in argon for samples of the BKBO system was carried out at $T \approx 700$ K for 1 h with subsequent cooling in argon to $T = 420$ K followed by quenching. In this case, a decrease in the intensity of the EPR signal was also observed, which again rose after an additional anneal in oxygen. Thus, these experiments show that the EPR signal in the half-field decreases as oxygen leaves the sample, resulting in the emergence of oxygen vacancies. To observe these signals, it is necessary to have high-quality samples with full oxygen stoichiometry. Loss of oxygen is pobably the reason for sample degradation and the disappearance of EPR signals under suboptimal handling of the samples.

We also carried out other experiments on the effect of heat treatment (temperature and duration of anneal, quench rate, etc.). Experiments on potassium and lead doping and a study of the effect of the argon and oxygen anneals, as well as other conditions of preparation, demonstrate convincingly that the EPR signals with $g_2 \approx 4.2$ are related to oxygen and are observed only in high-quality samples. The authors of Refs. 19, 23–25 who studied La_2CuO_4 and $La_{2-x}Sr_xCuO_4$ samples, also concluded that it is necessary to have high-quality samples without oxygen vacancies to observe EPR signals. However, it is more complicated to observe EPR signals in the LSCO system in the half-field due to the stronger signal of the Cu^{+2} ions. In addition, there are specific requirements on the local symmetry of the lattice when observing such EPR signals.[19]

4. DISCUSSION

The line with $g_2 \approx 4.2$, present in the EPR spectra of all the samples, is a distinguishing feature of the so-called ''$\Delta M_s = \pm 2$ forbidden'' transition between levels of the system with spin $S = 1$ (Ref. 21), where M_s is the magnetic quantum number.

The first observation of triplet EPR signals in HTSC materials and their interpretation belong to Thomann *et al.*[19] In $La_{2-x}Sr_xCuO_4$ samples without oxygen vacancies they were able to observe the corresponding pair of EPR signals. From the temperature dependence of the signal with $g \approx 4$ they were able to conclude that the interaction between the magnetic centers is ferromagnetic and that the ground state of this system is a triplet while its excited state is a singlet. Note that the presence of a Cu subsystem in the samples made it much harder to interpret the results. The authors of Ref. 19 could therefore not draw any definitive conclusions regarding the nature of the observed pairs. Nevertheless, they took localization of additional (doped) holes as the most probable reason for them.

In our opinion, such EPR signals can be understood on the basis of the model of ''ordered ionic–covalent bonds.''[14,26] In this model it is shown that in BKBO and BPBO systems, the insulating state results from charge density waves in the sublattice of oxygen and bismuth ions. For the sake of illustration, we can describe the same scheme in terms of different effective charges in the bismuth-ion sublattice and different effective charges in the oxygen-ion sublattice. This corresponds to the emergence of ionic–covalent Bi–O bonds with different degree of covalency. In other words, ionic–covalent $Bi–O^{-0.5}$ bonds with a higher fraction of covalency exist in the sample, along with the usual $Bi–O^{-2}$ bonds. A justification of the magnitudes of the effective charges of the Bi and O ions is given in Ref. 26. The ordering of such ionic–covalent bonds ($Bi–O^{-0.5}$) is superstructural, and leads to the formation of a dielectric gap. Due to overlap of the wave functions and the ordering of these bonds, a filled valence band without localized magnetic moments is formed.

Above we spoke of ideal, i.e., defect-free ordering of the $Bi–O^{-0.5}$ ionic–covalent bonds. This is diagrammed in Fig. 3a for a single Bi–O plane. In this case, as can be seen from the figure, the superstructural ordering of the $Bi–O^{-0.5}$ bonds leads to a doubling of the lattice period. The energy gap Σ in the vicinity of the point $(\pi/a)(1/2,1/2,/12)$ of the Brillouin zone is shown in Fig. 3b (here a is the distance between bismuth ions). In real samples, breaches arise in the ideal ordering of the $Bi–O^{-0.5}$ ionic–covalent bonds. One such breach is shown in Fig. 3c, where this breach, that is to say, ordering defect is encircled by a dashed line for clarity. It creates an acceptor impurity level in the band gap, denoted by the arrow in Fig. 3d.

This defect, which results from a breach in the ordering of the $O^{-0.5}$ ions, has a localized magnetic moment. For a sufficient concentration of such impurity centers, their magnetic moment can be detected by EPR. When a bound pair of such impurity centers forms, a singlet ground-state level and a triplet impurity excited level arise in the band gap in the vicinity of the neighboring bismuth ions. Such a bound pair of impurity centers is depicted in Fig. 3e and, for clarity, is also encircled by a dashed line. The energy spectrum of a crystal with such defects is shown in Fig. 3f. The single and triplet levels in the figure are represented by a corresponding pair of arrows.

Doping and the emergence of free carriers at the top of the valence band has a definite effect on the intensity of the EPR lines due to screening. However, the fact that EPR signals have been observed in the metallic phase of BKBO and BPBO suggests that their metallic state is indeed a state of a degenerate semiconductor. This conclusion is consistent with optical studies.[27–30] The conductivity and superconductivity in such a degenerate semiconductor are due to free carriers arising upon doping.

As is well known,[13] states at the top of the valence band are derived mainly from $2p$ oxygen states. The contribution of the bismuth states amounts to 3–5% and decreases with potassium doping (BKBO) or lead doping (BPBO). The contribution of potassium and lead to these states is essentially zero. Therefore, the state of the oxygen ''defect'' levels in the band gap Σ near the top of the valence band depends weakly on the potassium or lead concentration. When Bi is replaced by Pb (the replacement can be complete), $Bi–O^{-0.5}$

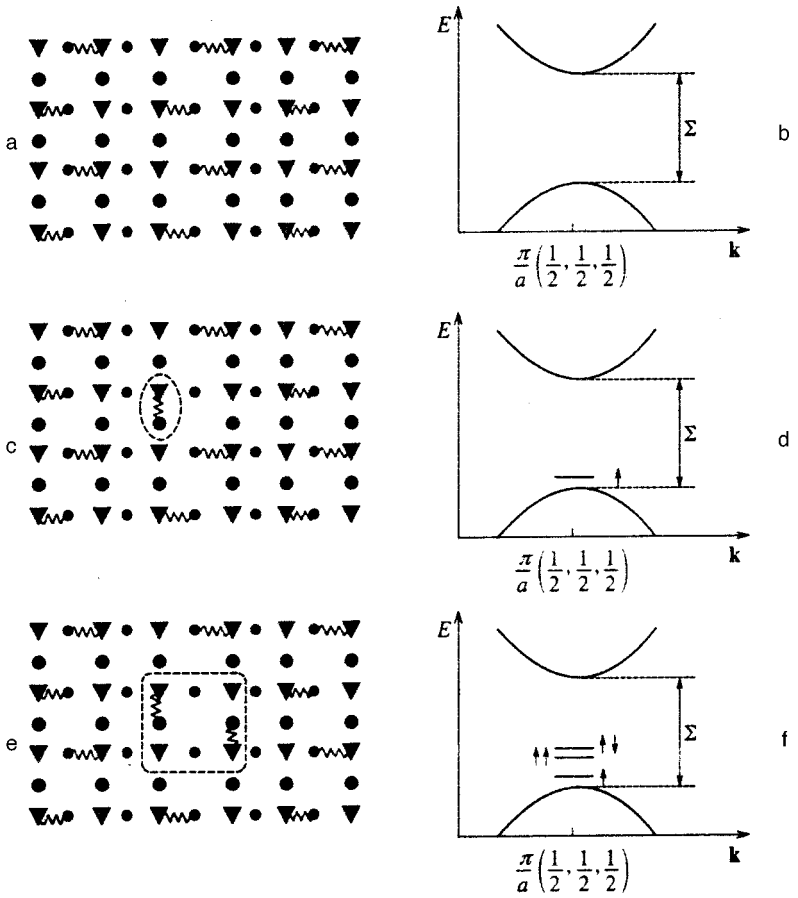

FIG. 3. Diagram of local violations of ideal ordering of the ionic–covalent bonds, also showing ''impurity'' levels in the band gap in the vicinity of the wave vector $\mathbf{k}=\pi/a$ (1/2,1/2,1/2); ▼—Bi ions; ●—O^{-2} ions; ▼⇝●—ionic–covalent bond ($Bi–O^{-0.5}$), Ba, K, and Pb ions not shown; a—diagram of ideal ordering of the ionic-covalent bonds for one Bi–O plane; b—dispersion $E(\mathbf{k})$ of the valence band and the conduction band with band gap Σ; c—defect in the ordering of the ionic–covalent bonds (encircled by the dashed line); d—''impurity'' level in the band gap with a local magnetic moment (represented by the arrow); e—bound pair of defects in the ordering of the ionic–covalent bonds for neighboring Bi ions (encircled by the dashed line); f—singlet (↑↓) and triplet (↑↑) levels in the band gap.

bonds are supplanted by corresponding ionic/covalent $Pb–O^{-0.5}$ bonds. This is the reason for the weak dependence of the EPR signals on the lead doping level.

The strong dependence of the EPR signals on the presence of oxygen vacancies is related to the fact that these are donor vacancies.[13,17,18] Compensation of acceptor levels thus arises, which leads to disappearance of the EPR signals.

Reference 20 reported the observation of EPR signals with $g_1\approx 2.1$ and $g_2\approx 4.2$ in some specimens of the BKBO system. The authors of Ref. 20 were able to observe such signals only in insulating samples with $x=0.13$ and $x=0.25$, whereas no EPR signals were observed in the parent compound $BaBiO_3$ or in superconducting samples with $x=0.40$ and $x=0.50$. They therefore concluded that localized hole pairs, being an EPR signal source, result exclusively from potassium doping. Since we were able to consistently observe EPR signals in undoped $BaBiO_3$, this conclusion must be characterized as erroneous. The lack of EPR signals in some samples in Ref. 20 may be linked to ''aging effects'' (the authors themselves noted this effect in their work).

We now consider the temperature dependence of the observed EPR signal with $g_2\approx 4.2$ in more detail. Since the position, shape, and width of this EPR line did not change with temperature over the range investigated, the temperature dependence of the intensity of the EPR signal I_p can be analyzed using the formula[21]

$$I_p=\frac{C}{T+\Theta}\exp\left(-\frac{J_p}{T}\right), \tag{1}$$

where J_p is the exchange coupling constant inside the pair of localized magnetic moments, Θ is the paramagnetic Curie temperature characterizing the interaction in the system of triplet pairs, and C is a constant. Analysis of the dependence (1) for the samples investigated showed that $J_p\geqslant 0$. For example, for $BaBiO_3$ we have $J_p=0.9\pm 0.5$ K, and for $Ba_{0.6}K_{0.4}BiO_3$ we obtain $J_p=3.9\pm 0.5$ K. Similar values are also observed for other compositions. Taking into account the smallness of J_p, we can replace formula (1) by the simplified dependence

$$I_p\approx\frac{C}{T+\Theta}, \tag{2}$$

which does not include the activation factor $\exp(-J_p/T)$.

The dependence $I_p^{-1}(T)$ is plotted in Fig. 2, where the solid line corresponds to formula (2). Least-squares fitting of the experimental data plotted in Fig. 2 yields $\Theta=4\pm 2$ K for the investigated samples, i.e., the quantity $\Theta>0$, which indicates an antiferromagnetic interaction between the triplet localized pairs. It is clear from Fig. 2 that samples of differing composition have very similar values of Θ.

From the above estimate of $J_p>0$, it follows[21,31] that the ground state of a localized pair is a singlet, and the excited state is a triplet, as shown in Fig. 3f for acceptor impurity levels.

5. CONCLUSION

To sum up, we have shown that the observed EPR signals are due to oxygen ions and are detected only in high-quality samples. The presence of magnetic moments on the localized pairs probably indicates that aside from the $Bi–O^{-2}$ ionic bonds, bonds of type $Bi–O^{-\alpha}$ (or $Pb–O^{-\alpha}$ in $BaPbO_3$) also exist, where $\alpha<2$, i.e., ionic–covalent bonds exist between bismuth with oxygen. Their observation by EPR is possible in the case of a breach of ideal ordering. It follows from these experiments that besides the familiar superstructural ordering in the bismuth-ion sublattice, an additional superstructural ordering (or charge density wave) probably also exists in the oxygen-ion sublattice. The presence of a charge density wave in the oxygen-ion sublattice should affect the band structure of oxide HTSC systems.

We thank A. V. Tsikunov for performing the elemental analysis of the samples on the LAMMA-1000 system and S. V. Verkhovskiĭ for discussion of the results. This work was supported by the Scientific Committee on Current Directions in the Physics of Condensed Media (Project 96086) and the Russian Fund for Fundamental Research (Project No. 98-02-17411).

[1] M. Braden, W. Reichardt, W. Schmidbauer, A. S. Ivanov, and A. Yu. Rumiantsev, J. Supercond. **8**, 595 (1995).
[2] W. Reichardt, L. Pintschovius, N. Pyka, P. Schweiß, A. Erb, P. Bourges, G. Collin, J. Rossat-Mignod, I. Y. Henry, A. S. Ivanov, N. L. Mitrofanov, and A. Yu. Rumiantsev, J. Supercond. **7**, 399 (1994).
[3] H. You, U. Welp, and Y. Fang, Phys. Rev. B **43**, 3660 (1991).
[4] N. V. Anshukova, A. I. Golovashkin, L. I. Ivanova, I. B. Krinetskii, K. V. Kraiskaya, L. I. Leonjuk, and A. P. Rusakov, Physica C **282–287**, 1065 (1997).
[5] V. F. Gantmakher, L. A. Klinkova, N. V. Barkovskii, G. E. Tsydynzhapov, S. Wiegers, and A. K. Geim, Phys. Rev. B **54**, 6133 (1996).
[6] N. V. Anshukova, A. I. Golovashkin, L. I. Ivanova, and A. P. Rusakov, Usp. Fiz. Nauk **167**, 887 (1997).
[7] W. D. Mosley, J. W. Dykes, R. N. Shelton, P. A. Sterne, and R. H. Howell, Phys. Rev. Lett. **74**, 1271 (1994).
[8] D. M. King, Z.-X. Shen, D. S. Dessau, B. O. Wells, W. E. Spicer, A. J. Arko, D. S. Marshall, J. DiCarlo, A. G. Loeser, C. H. Park, E. R. Ratner, J. L. Peng, Z. Y. Li, and R. L. Greene, Phys. Rev. Lett. **70**, 3159 (1993).
[9] M. C. Schabel, C.-H. Park, A. Matsuura, Z. X. Shen, D. A. Bonn, Ruixing Liang, and W. N. Hardy, Phys. Rev. B **55**, 2796 (1997).
[10] H. Ding, T. Yokoya, J. C. Campuzano, T. Takahashi, M. Randeria, M. R. Norman, T. Mochiku, K. Kadowaki, and J. Giapintzakis, Nature (London) **382**, 51 (1996).
[11] J. M. Harris, P. J. White, Z.-X. Shen, H. Ikeda, R. Yoshizaki, H. Eisaki, S. Uchida, W. D. Si, J. W. Xiong, Z.-X. Zhao, and D. S. Dessau, Phys. Rev. Lett. **79**, 143 (1997).
[12] Z.-X. Shen, W. E. Spicer, D. M. King, D. S. Dessau, and B. O. Wells, Science **267**, 343 (1995).
[13] N. V. Anshukova, A. I. Golovashkin, L. I. Ivanova, O. T. Malyuchkov, A. P. Rusakov, Zh. Éksp. Teor. Fiz. **108**, 2132 (1995) [JETP **81**, 1163 (1995)].
[14] N. V. Anshukova, A. I. Golovashkin, L. I. Ivanova, O. T. Maljuchkov, and A. P. Rusakov, Physica C **273**, 151 (1996).
[15] A. Yakubovskii, S. Gudenko, A. Rusakov, A. Golovashkin, and S. Verkhovskii, Physica C **282–287**, 1929 (1997).
[16] Shiyou Pei, J. Jorgensen, B. Dabrowski, D. G. Hinks, D. R. Richards, A. W. Mitchell, J. M. Newsam, S. K. Sinha, D. Vaknin, and A. J. Jacobson, Phys. Rev. B **41**, 4126 (1990).
[17] T. Hashimoto and H. Kawazoe, Solid State Commun. **87**, 251 (1993).
[18] T. Hashimoto, H. Kawazoe, and H. Shimamura, Physica C **223**, 131 (1994).
[19] H. Thomann, R. A. Klemm, D. C. Johnston, P. J. Tindall, H. Jin, and D. P. Goshorn, Phys. Rev. B **38**, 6552 (1988).
[20] S. K. Misra, S. I. Andronenko, R. R. Andronenko, and L. P. Mezentseva, Phys. Rev. B **53**, 9442 (1996).
[21] J. E. Wertz and J. R. Bolton, *Electron Spin Resonance: Elementary Theory and Practical Applications*, McGraw-Hill, New York (1972).
[22] G. Kh. Panova, M. A. Shikov, M. N. Khlopkin, A. P. Zhernov, N. V. Anshukova, A. I. Golovashkin, L. I. Ivanova, and A. P. Rusakov, Zh. Éksp. Teor. Fiz. **103**, 605 (1993) [JETP **76**, 302 (1993)].
[23] B. I. Kochelaev, J. Sichelschmidt, B. Elschner, W. Lemor, and A. Loidl, Phys. Rev. Lett. **79**, 4274 (1997).
[24] P. Simon, J. M. Bassat, S. B. Oseroff, Z. Fisk, S.-W. Cheong, A. Wattiaux, and Sheldon Schultz, Phys. Rev. B **48**, 4216 (1993).
[25] A. Punnoose and R. J. Singh, Int. J. Mod. Phys. B **9**, 1123 (1995).
[26] N. V. Anshukova, A. I. Golovashkin, L. I. Ivanova, and A. P. Rusakov, *Brief Reports on Physics* [in Russian], FIAN No. 8, 19 (1998).
[27] S. Uchida, K. Kirazawa, and S. Tanaka, Phase Transit. **8**, 95 (1987).
[28] S. H. Blanton, R. T. Collins, K. H. Kelleher, L. D. Rotter, Z. Schlesinger, D. G. Hinks, and Y. Zheng, Phys. Rev. B **47**, 996 (1993).
[29] M. A. Karlow, S. L. Cooper, A. L. Kotz, M. V. Klein, P. D. Han, and D. A. Payne, Phys. Rev. B **48**, 6499 (1993).
[30] A. V. Puchkov, T. Timusk, M. A. Karlow, S. L. Cooper, P. D. Han, and D. A. Payne, Phys. Rev. B **54**, 6686 (1996).
[31] Yu. V. Yablokov, V. K. Voronkova, and L. V. Mosina, *Paramagnetic Resonance of Exchange Clusters* [in Russian], Nauka, Moscow (1988).

Translated by Paul F. Schippnick

Donor states in tunnel-coupled quantum wells

F. T. Vas'ko*) and V. I. Pipa

Institute of Semiconductor Physics, National Academy of Sciences, 252650 Kiev, Ukraine
(Submitted 12 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1337–1352 (April 1999)

We study the energy spectrum of the impurity states in tunnel-coupled double quantum wells for Coulomb and short-range donor potentials. We calculate the impurity contribution and the density of states and detect the transformation of a localized donor state into a resonant state when the binding energy of the donor in an isolated quantum well is less than the separation of the energy levels of the double quantum wells. In the opposite case, where the binding energy is greater than the level separation, there is tunneling repulsion between adjacent impurity levels, with the degree of degeneracy of the levels changing when there is tunneling mixing of the ground and excited impurity states from different wells. Resonant states emerge in an asymmetric double quantum well, while in a symmetric double quantum well the impurity level at the barrier's center proves to be localized even against the background of the continuum. The calculations are based on a general expression for the impurity contribution to the density of states in terms of a 2-by-2 matrix Green's function, i.e., only a pair of tunnel-coupled levels of the double quantum wells is taken into account. For an impurity with a short-range potential, we derive a matrix generalization of the Koster–Slater solution, while the impurity with a Coulomb potential is analyzed by using the approximation of a narrow resonance and close arrangement of the repulsive levels. © *1999 American Institute of Physics.*
[S1063-7761(99)01304-9]

1. INTRODUCTION

The transport and optical properties of double quantum wells change significantly (see Refs. 1–7 and the literature cited therein) due to tunneling mixing of the electronic states of the left- (l) and right-hand (r) quantum wells. The structure of the donor states in double quantum wells can also be significantly altered in comparison to ordinary bulk[8] and two-dimensional[9] states. This modification of the donor states is reflected in qualitative features (see Fig. 1) when the binding energy is comparable to the level separation in the double quantum wells, Δ_T, whose value is determined by the height and width of the barrier (the upper band diagrams in Fig. 1 correspond to weak interwell tunneling, while the lower diagrams demonstrate the tunneling mixing effect). If the binding energy of a Coulomb impurity in the l quantum well is higher than the energy level separation Δ_T in the absence of tunneling, at certain values of Δ the ground-state level in the l quantum well may coincide with the ground or excited state in the r quantum well (see the upper band diagrams in Figs. 1a and 1b). Due to tunneling mixing, these levels repel each other near their crossing point (the anticrossing effect), as shown in the lower band diagrams in Figs. 1a and 1b. When the ground level mixes with an excited degenerate state (Fig. 1b), the degree of degeneracy of the emerging state is lower. But if the energy of the electron of the donor state in the l quantum well shows up against the background of the continuum of the r quantum well (see upper diagram in Fig. 1c), then due to tunneling mixing this state becomes resonant (see the lower band diagram in Fig. 1c).

Resonant donor states are studied in this paper for the cases of short-range defects and Coulomb impurities. These two types of impurity have different energy spectra even when there is only one quantum well: a point defect produces a single bound state, while a Coulomb impurity produces a series of levels (which leads to the possibility of changing the degree of degeneracy by level mixing; see Fig. 1b). When the impurity is at the center of the barrier of a symmetric double quantum well, the impurity potentials in the l and r quantum wells coincide, with the result that longitudinal electron localization and the tunneling mixing of size-quantized states prove to be independent. In this case, a localized, i.e., nonresonant, state emerges against the background of the continuum.

The foregoing modification of the impurity energy spectrum has a powerful effect on electron transport and the optical properties of a double quantum well at moderate doping, but thus far only heavily doped and pure double quantum wells have been thoroughly studied.[1,5] Tunneling mixing of the ground and excited donor states in a double quantum well was detected by Ranganathan *et al.*,[10] who used the method of far-infrared radiation magnetotransmission (Dzyubenko and Yablonskiĭ[11] also examined the mixing of the ground and excited magnetoexcitonic states). These results were discussed on the basis of variational calculations of donor binding energy, while the resonant states and the features of the spectrum of excited states we have just discussed cannot be obtained by the standard variational approach. We also note that the special features of the excitonic energy spectrum are similar to those of a Coulomb impurity. Although the ground state of excitons in a double quantum

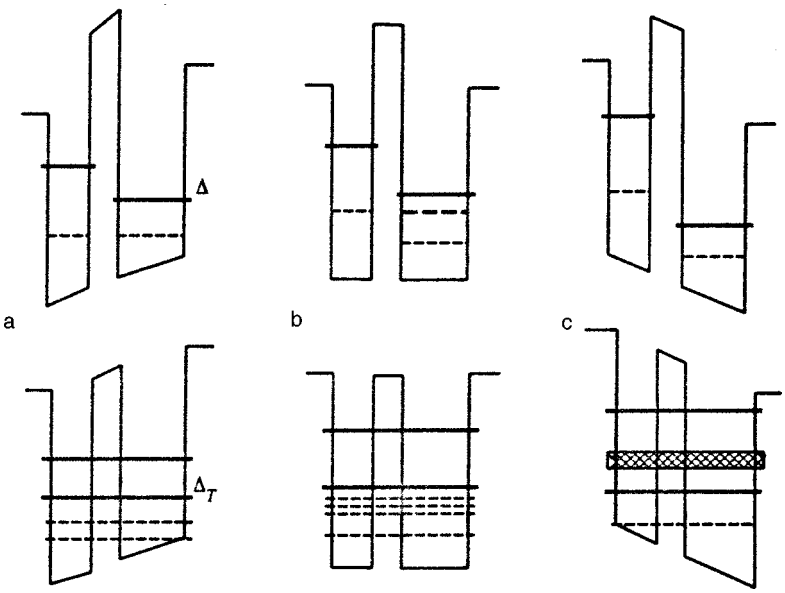

FIG. 1. Band diagram, energies of the subband extrema (solid lines), and impurity levels (dashed lines) for double quantum wells in the absence of tunneling coupling (upper panel) and for tunnel-coupled quantum wells (lower panel). a) Repulsion of ground states; b) mixing of the excited and ground states from the r and l quantum wells; c) transformation of a localized state into a resonant state.

well has been thoroughly studied (see Ref. 12 and the literature cited therein), the appreciable broadening of excitonic lines due to tunneling is discussed only by Fox *et al.*[13] and Oberli *et al.*[14] The presence of such features in excitonic absorption spectra is corroborated by the numerical calculations of Glutsch *et al.*[15] Crossing of localized levels was studied only for magnetoexcitonic states.[16]

In this paper we calculate the impurity contribution to the density of states by employing the one-electron Green's function[17] in the 2-by-2 matrix representation, which allows only for a pair of the lowest tunnel-coupled levels in the l and r quantum wells. We obtain an exact solution of the Koster–Slater type for a short-range potential, while the Coulomb potential is analyzed by using the approximations of a narrow resonance and close tunnel-coupled levels.

In Sec. 2 we derive an expression for the impurity contribution to the density of states. In Secs. 3 and 4 we apply this formalism to point defects and Coulomb impurities. Finally, in Sec. 5 we discuss the results and draw conclusions.

2. BASIC EQUATIONS

We begin with the formalism describing the impurity contribution to the density of states $\rho(E)$ of the double quantum wells. This contribution can be expressed in terms of the retarded Green's function G_ε in the usual way:

$$\rho(E)=\lim_{\varepsilon\to E+i0}\ \mathrm{Im}\,\frac{2}{\pi}\int dz\sum_{\mathbf{p}}\ G_\varepsilon(\mathbf{p}z,\mathbf{p}z). \tag{1}$$

Here we have used the $\mathbf{p},z$-representation, $\mathbf{p}$ is the 2D momentum of the electron, the z axis is perpendicular to the plane of the 2D layer, and the normalization area is taken to be unity. The Green's function satisfies the equation

$$(H_{\mathrm{DQW}}-\varepsilon)G_\varepsilon(\mathbf{p}z,\mathbf{p}'z')$$

$$-\sum_{\mathbf{p}_1}\ V(|\mathbf{p}-\mathbf{p}_1|,z)G_\varepsilon(\mathbf{p}_1z,\mathbf{p}'z')=\delta_{\mathbf{pp}'}\delta(z-z'), \tag{2}$$

where H_{DQW} is the double-quantum-well Hamiltonian in the absence of impurities. The potential energy of the Coulomb center in (2) is given by

$$V(p,z)=\frac{2\pi e^2\hbar}{\kappa p}\exp\left(-\frac{p|z-z_D|}{\hbar}\right), \tag{3}$$

in which the dielectric constant κ is homogeneous in the direction perpendicular to the double quantum well; the Coulomb center is at $(0,0,z_D)$. For a point defect (i.e., a substitutional impurity),

$$V(p,z)=u_p\delta(z-z_D),$$

where the delta function is localized on a scale of the order of the lattice constant a, and the potential u_p is constant for $p<p_m$ and small for $p>p_m$ (here the maximum momentum p_m is of order $\hbar/a$).

Shallow impurity states, for which the binding energy is small compared to the distance between the levels in the quantum well, can be described by allowing for only the lowest tunnel-coupled double-quantum-well states. For such states the Green's function can be expanded in the orbitals of the l and r quantum wells, the latter being denoted by $\varphi_l(z)$ and $\varphi_r(z)$:

$$G_\varepsilon(\mathbf{p}z,\mathbf{p}'z')=\sum_{jj'}\ \varphi_j(z)G_\varepsilon(\mathbf{p}j,\mathbf{p}'j')\varphi_{j'}(z'). \tag{4}$$

The expansion coefficients $G_\varepsilon(\mathbf{p}j,\mathbf{p}'j')$ for a 2-by-2 matrix Green's function $\hat{G}_\varepsilon(\mathbf{p},\mathbf{p}')$, which in such an ''isospin'' representation is determined by the equation

$$(\varepsilon_p+\hat{h}-\varepsilon)\hat{G}_\varepsilon(\mathbf{p},\mathbf{p}')$$

$$-\sum_{\mathbf{p}_1}\ \hat{V}(|\mathbf{p}-\mathbf{p}_1|)\hat{G}_\varepsilon(\mathbf{p}_1,\mathbf{p}')=\delta_{\mathbf{pp}'}, \tag{5}$$

where $\varepsilon_p=p^2/2m$ is the kinetic energy, $\hat{h}=(\Delta/2)\hat{\sigma}_z+T\hat{\sigma}_x$ is the Hamiltonian matrix describing transverse motion, the $\hat{\sigma}_i$ are the Pauli matrices, and Δ is the distance between the

lowest levels in the isolated l and r quantum wells. The expressions for Δ and the tunneling matrix element T are given in Ref. 18 for the flat-band model. We assume that the level separation Δ and the binding energy of an electron on a donor are small compared to the distance to the higher levels of the double quantum wells. In this case, the tunneling-Hamiltonian approximation accounts exactly for the intermingling of the pair of lowest states in the double quantum wells, with only the contributions of the higher levels ignored. Only the diagonal elements of the impurity-potential matrix $\hat{V}(p)$ in Eq. (5) are important (the off-diagonal elements are small compared to $T\hat{\sigma}_x$). These elements are given by

$$V_j(p)=\frac{2\pi e^2\hbar}{\kappa p}\int_{-d_j/2}^{d_j/2}dz\,\varphi_j^2(z)\exp\left(-\frac{p|z-z_{Dj}|}{\hbar}\right), \quad (6)$$

where the z-coordinate is measured from the center of the jth quantum well (z_j), $z_{Dj}=(z_D-z_j)$, and d_j is the width of the jth quantum well. For a point defect localized in the jth quantum well we use $V_j(p)=u_p\varphi_j^2(z_D)$, ignoring the exponentially small off-diagonal matrix elements $u_p\varphi_l(z_D)\varphi_r(z_D)$. An expression similar to that for $V_j(p)$ can also be written for a potential generated by a small-scale inhomogeneity of the heteroboundaries of double quantum wells (such a potential was described in Ref. 19 in connection with the problem of scattering by nonideal heteroboundaries). In this case,

$$V_j(p)=\frac{2\pi\varepsilon_j\xi b^2}{d_j}, \quad (7)$$

where ε_j is the level energy in the jth quantum well, and ξ and b are the height and longitudinal size of the inhomogeneity.

The density of states (1) in the ''isospin'' representation can be transformed to

$$\rho(E)=\lim_{\varepsilon\to E+i0}\mathrm{Im}\,\frac{2}{\pi}\,\mathrm{Tr}\sum_{\mathbf{p}}\hat{G}_\varepsilon(\mathbf{p},\mathbf{p}), \quad (8)$$

where Tr denotes the trace, the sum of the diagonal matrix elements. Thus, to describe both localized and resonant impurity states we must solve the matrix integral equation (5) and do the summation in (8). Below we carry out these calculations analytically for point defects described by the matrix form of the Koster–Slater equation; for Coulomb donors we use additional approximations.

3. POINT DEFECTS

By introducing the retarded Green's function in the absence of impurities, $\hat{g}_\varepsilon(p)\delta_{\mathbf{pp}'}$, where

$$\hat{g}_\varepsilon(p)=(\varepsilon_p+\hat{h}-\varepsilon)^{-1}, \quad (9)$$

we can write Eq. (5) as

$$\hat{G}_\varepsilon(\mathbf{p},\mathbf{p}')=\hat{g}_\varepsilon(p)\delta_{\mathbf{pp}'}+\hat{g}_\varepsilon(p)\hat{V}\sum_{\mathbf{p}_1}{}'\,\hat{G}_\varepsilon(\mathbf{p}_1,\mathbf{p}'). \quad (10)$$

Here the sum $\Sigma'_{\mathbf{p}_1}$ is calculated for $|\mathbf{p}_1|<P_m$, and the potential matrix $\hat{V}$ is determined by the components (7), which are $\mathbf{p}$-independent. If we ignore small subbarrier penetration, then for an impurity localized in the jth well we have $\hat{V}=V_j\hat{P}_j$, where $\hat{P}_l=(1+\hat{\sigma}_z)/2$ and $\hat{P}_r=(1-\hat{\sigma}_z)/2$ are the projection operators on the orbitals of the l and r quantum wells, and V_j has been defined in (7).

To solve Eq. (10), we do the summation over $\mathbf{p}$ and write the sum as follows:

$$\sum_{\mathbf{p}}{}'\,\hat{G}_\varepsilon(\mathbf{p},\mathbf{p}')=\left[1-\sum_{\mathbf{p}}{}'\,\hat{g}_\varepsilon(p)\hat{V}\right]^{-1}\hat{g}_\varepsilon(p'). \quad (11)$$

Substituting this into the right-hand side of Eq. (10) yields

$$\hat{G}_\varepsilon(\mathbf{p},\mathbf{p}')=\hat{g}_\varepsilon(p)\left\{\delta_{\mathbf{pp}'}+\hat{V}\left[1-\sum_{\mathbf{p}}{}'\,\hat{g}_\varepsilon(p)\hat{V}\right]^{-1}\hat{g}_\varepsilon(p')\right\}, \quad (12)$$

where the second term describes the perturbation introduced by the impurity. Inserting (12) into (8), we obtain an expression for the impurity contribution to the density of states:

$$\delta\rho_{\mathrm{imp}}(E)=\lim_{\varepsilon\to E+i0}\mathrm{Im}\,\frac{2n_{\mathrm{imp}}}{\pi}\,\mathrm{Tr}\sum_{\mathbf{p}}{}'\,\hat{g}_\varepsilon(p)\hat{V}\times\left[1-\sum_{\mathbf{p}_1}{}'\,\hat{g}_\varepsilon(p_1)\hat{V}\right]^{-1}\hat{g}_\varepsilon(p). \quad (13)$$

The introduction of the impurity concentration n_{imp} into (13) presupposes that the electronic states at different impurities do not overlap (i.e., the distance between impurities in the plane of a double quantum well does not exceed the radii of the states localized at the impurities).

Summation over $\mathbf{p}_1$ in Eq. (13) yields the matrix

$$\hat{\lambda}(\varepsilon)=\frac{\rho_{2\mathrm{D}}}{2}\left[\hat{V}I_+(\varepsilon)+\left(\frac{\Delta}{\Delta_T}\hat{\sigma}_z+\frac{2T}{\Delta_T}\hat{\sigma}_x\right)\hat{V}I_-(\varepsilon)\right], \quad (14)$$

where

$$I_+(\varepsilon)=\ln\frac{\xi_m^2}{(\varepsilon+\Delta_T/2)(\varepsilon-\Delta_T/2)},$$

$$I_-(\varepsilon)=\ln\frac{\varepsilon+\Delta_T/2}{\varepsilon-\Delta_T/2}. \quad (15)$$

Here $\rho_{2\mathrm{D}}=m/\pi\hbar^2$, $\xi_m=p_m^2/2m$, and $\Delta_T=\sqrt{\Delta^2+4T^2}$ is the distance between the tunnel-coupled levels (the energies $\pm\Delta_T/2$ determine the positions of the extrema of the tunnel-coupled subbands). Using (14), we can write

$$\delta\rho_{\mathrm{imp}}(E)=\lim_{\varepsilon\to E+i0}\mathrm{Im}\,\frac{2n_{\mathrm{imp}}}{\pi}\,\mathrm{Tr}[1-\hat{\lambda}(\varepsilon)]^{-1}\frac{d\hat{\lambda}(\varepsilon)}{d\varepsilon}. \quad (16)$$

After calculating the trace, we obtain the final analytic expression,

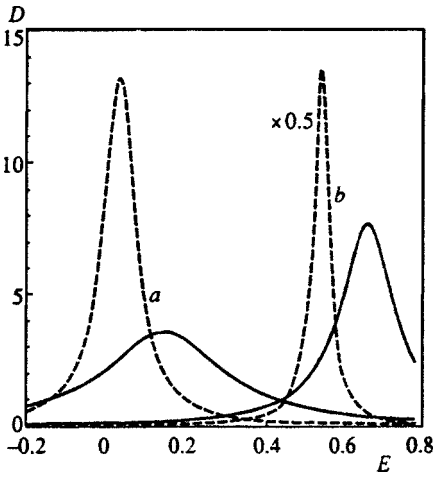

FIG. 2. Contribution of a short-range defect to the density of states $D(E)=\delta\rho_{\rm imp}(E)/2n_{\rm imp}$ for double quantum wells with $\Delta=2$ meV for various values of T: $T=0.5$ meV (solid curves), and $T=0.25$ meV (dashed curves). The energy E is measured in meV, and $D(E)$ is measured in in meV^{-1}. The donor level in an isolated l quantum well (at $T=0$) is fixed at $E_0=0$ meV (a) and $E_0=0.5$ meV (b).

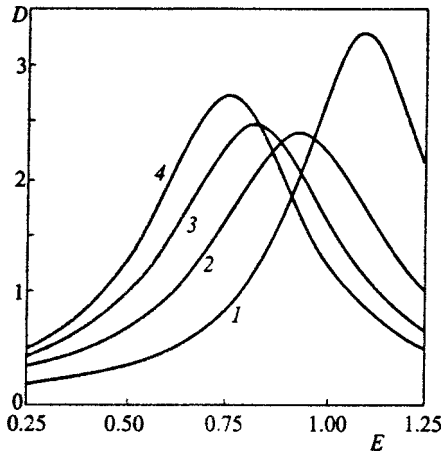

FIG. 3. Same as in Fig. 2 for $T=1$ meV and various level separations: curve *1*, $\Delta=2$ meV; curve *2*, $\Delta=3$ meV; curve *3*, $\Delta=4$ meV; curve *4*, $\Delta=5$ meV.

$$\delta\rho_{\rm imp}(E)=-\lim_{\varepsilon\to E+i0}{\rm Im}\,\frac{2n_{\rm imp}}{\pi}\frac{d}{d\varepsilon}\ln\left[1-\frac{\rho_{\rm 2D}V_j}{4}J(\varepsilon)\right], \tag{17}$$

where $J(\varepsilon)=I_+(\varepsilon)\pm(\Delta/\Delta_T)I_-(\varepsilon)$, and the plus and minus signs refer to impurities localized in the l and r quantum wells, respectively.

Suppose that the impurity is in the l well. Direct calculation of (17) for $E<-\Delta_T/2$ yields $\delta\rho_{\rm imp}(E)=2n_{\rm imp}\delta(E-E_L)$, where the energy E_L of the localized state can be found by solving the equation

$$\left(1-\frac{\Delta}{\Delta_T}\right)\ln\left|\frac{E+\Delta_T/2}{E_0-\Delta/2}\right|+\left(1+\frac{\Delta}{\Delta_T}\right)\ln\left|\frac{E-\Delta_T/2}{E_0-\Delta/2}\right|=0, \tag{18}$$

where E_0 is the energy of a localized state in an isolated l well. In a symmetric DQW, i.e., $\Delta=0$, the energy E_L is $-\sqrt{E_0^2+T^2}$, and as $|\Delta|$ increases, the energy approaches the edge of the continuum.

But if the solution of Eq. (18) is found for $|E|<\Delta_T/2$, then due to tunneling mixing of the localized level and the states of the continuum, a resonant state is formed. First we give the simple analytic expressions for weak tunnel coupling, $T/\Delta^2\ll 1$, assuming that the level E_0 is far from the band edge, i.e., $(T/\Delta)^2\ln[(\Delta/2+E_0)/(\Delta/2-E_0)]\ll 1$. We denote the solution of Eq. (18) for this case by E_R. For the impurity contribution to the density of states near E_R (where $|E-E_R|\ll|E_R\pm\Delta/2|$) we have

$$\delta\rho_{\rm imp}(E)=\frac{2n_{\rm imp}}{\pi}\frac{\Gamma}{(E-E_R)^2+\Gamma^2}, \tag{19}$$

where the energy $\Gamma=\pi(T/\Delta)^2(\Delta/2-E_0)$ is the level's half-width. Equation (19) is valid for a narrow resonance, with Γ small compared to E_R. The results of calculating the impurity contribution $\delta\rho_{\rm imp}(E)$ to the density of states [using the general expression (17)] are depicted in Figs. 2 and 3. Figure 2 demonstrates that as the interwell tunnel coupling increases, the peak decreases, broadens, and is shifted toward higher energies. For the limit of uncoupled quantum wells ($T=0$), we have a delta-function peak at energy E_0. Comparison of the values of $\delta\rho_{\rm imp}(E)$ obtained for various values of E_0 shows that the deeper the level, the more effective its transformation into a resonant state (i.e., the level width proves to be greater, all other parameters being equal). The shape of the peak in $\delta\rho_{\rm imp}(E)$ calculated for various values of Δ is depicted in Fig. 3. Clearly, as the level separation increases, the peak is shifted toward lower energies, with its amplitude varying nonmonotonically. As Δ decreases, i.e., a symmetric double quantum well is realized, the peak is shifted toward the edge of the continuum of the upper subbands and broadens. If $\Delta/T<(\pi-2)/\sqrt{\pi-1}$ [this inequality follows from (17)], the impurity contribution to the density of states monotonically increases with energy, i.e., there is no resonant state.

4. TUNNELING MODIFICATION OF COULOMB STATES

To describe a Coulomb donor, we use the representation of the diagonal Hamiltonian $\hat S^{-1}\hat H\hat S$, where $\hat S=\exp(i\psi\hat\sigma_y)$, with the angle ψ determined by $\tan 2\psi=2T/\Delta_T$ (such a description of electrons in a double quantum well was introduced in Ref. 18). The integral equation for the Green's function in such a representation, $\hat{\mathscr{G}}_\varepsilon(\mathbf{p},\mathbf{p}')=\hat S^{-1}\hat G_\varepsilon(\mathbf{p},\mathbf{p}')\hat S$, can be written

$$\left(\varepsilon_p+\frac{\Delta_T}{2}\hat\sigma_z-\varepsilon\right)\hat{\mathscr{G}}_\varepsilon(\mathbf{p},\mathbf{p}')-\sum_{\mathbf{p}_1}\hat{\mathscr{V}}(|\mathbf{p}-\mathbf{p}_1|)\hat{\mathscr{G}}_\varepsilon(\mathbf{p}_1,\mathbf{p}')=\delta_{\mathbf{pp}'}. \tag{20}$$

Here $\hat{\mathscr{V}}(p)=\hat S^{-1}\hat V(p)\hat S$, where the diagonal matrix $\hat V(p)$ is defined by its components (6). At this point in our discourse it is convenient to introduce an auxiliary Green's function $\hat g_\varepsilon(\mathbf{p},\mathbf{p}')$ satisfying Eq. (20), which takes into account only the diagonal part of the matrix $\hat{\mathscr{V}}$. The matrix $\hat g_\varepsilon(\mathbf{p},\mathbf{p}')$ is diagonal, and its components are

$$g_{\varepsilon}^{\pm}(\mathbf{p},\mathbf{p}')=\sum_{\lambda}\frac{\phi_{\lambda}^{\pm}(\mathbf{p})\phi_{\lambda}^{\pm}(\mathbf{p}')^{*}}{E_{\lambda}^{\pm}-\varepsilon}, \tag{21}$$

where the wave functions $\phi_{\lambda}^{\pm}$ and the eigenvalues $E_{\lambda}^{\pm}$ can be found by solving the equation

$$\left(\varepsilon_p\pm\frac{\Delta_T}{2}-E_{\lambda}^{\pm}\right)\phi_{\lambda}^{\pm}(\mathbf{p})-\sum_{\mathbf{p}_1}V_{\pm}(|\mathbf{p}-\mathbf{p}_1|)\phi_{\lambda}^{\pm}(\mathbf{p}_1)=0, \tag{22}$$

$$2V_{\pm}(p)=[V_l(p)+V_r(p)]\pm\frac{\Delta}{\Delta_T}[V_l(p)-V_r(p)].$$

The $\pm$ states correspond to localized donor states related to the upper $(+)$ or lower $(-)$ subbands of the double quantum wells without allowance for Coulomb mixing. The off-diagonal elements of the potential matrix $\hat{\mathscr{V}}$ are given by $w(p)\hat{\sigma}_x$, where

$$w(p)=\frac{T}{\Delta_T}[V_l(p)-V_r(p)]. \tag{23}$$

Using the diagonal matrix $\hat{g}_{\varepsilon}(\mathbf{p},\mathbf{p}')$, we rewrite Eq. (20):

$$\hat{\mathscr{G}}_{\varepsilon}(\mathbf{p},\mathbf{p}')=\hat{g}_{\varepsilon}(\mathbf{p},\mathbf{p}')-\sum_{\mathbf{p}_1,\mathbf{p}_2}\hat{g}_{\varepsilon}(\mathbf{p},\mathbf{p}_1)w(|\mathbf{p}_1-\mathbf{p}_2|)\times\hat{\sigma}_x\hat{\mathscr{G}}_{\varepsilon}(\mathbf{p}_2,\mathbf{p}'). \tag{24}$$

Eliminating the off-diagonal components of $\hat{\mathscr{G}}_{\varepsilon}(\mathbf{p},\mathbf{p}')$ from the system of integral equations (24), we obtain two independent integral equations for the diagonal components $\mathscr{G}_{\varepsilon}^{+}$ and $\mathscr{G}_{\varepsilon}^{-}$, which describe donor states related to the upper and lower subbands, respectively. We write these equations for the Green's functions $\mathscr{G}_{\varepsilon}^{\pm}(\lambda,\lambda')$ in the λ-representation, which is introduced by the following relationship:

$$\mathscr{G}_{\varepsilon}^{\pm}(\mathbf{p},\mathbf{p}')=\sum_{\lambda,\lambda'}\phi_{\lambda}^{\pm}(\mathbf{p})\mathscr{G}_{\varepsilon}^{\pm}(\lambda,\lambda')\phi_{\lambda'}^{\pm}(\mathbf{p}')^{*}. \tag{25}$$

The system of equations for $\mathscr{G}_{\varepsilon}^{\pm}(\lambda,\lambda')$ has the form

$$(E_{\lambda}^{\pm}-\varepsilon)\mathscr{G}_{\varepsilon}^{\pm}(\lambda,\lambda')=\delta_{\lambda\lambda'}+\sum_{\lambda_1}W_{\varepsilon}^{\mp}(\lambda,\lambda_1)\mathscr{G}_{\varepsilon}^{\pm}(\lambda_1,\lambda'), \tag{26}$$

where the kernel $W_{\varepsilon}^{\mp}$ is given by

$$W_{\varepsilon}^{\mp}(\lambda,\lambda')=\sum_{\mathbf{p}_1,\mathbf{p}_1',\mathbf{p}_2,\mathbf{p}_2'}\phi_{\lambda}^{\pm}(\mathbf{p}_1)^{*}w(|\mathbf{p}_1-\mathbf{p}_1'|)\times g_{\varepsilon}^{\mp}(\mathbf{p}_1',\mathbf{p}_2')w(|\mathbf{p}_2'-\mathbf{p}_2|)\phi_{\lambda'}^{\pm}(\mathbf{p}_2). \tag{27}$$

Note that Eqs. (26) and (27) are exact if one uses the exact eigenfunctions determined by Eq. (22).

In the simplest case of a symmetric double quantum well with an impurity at the barrier's center, Eq. (6) yields $V_l(p)=V_r(p)$, so that $w(p)=0$. As a result, even when the tunnel coupling of the l and r quantum wells is strong, $\delta\rho_{\mathrm{imp}}(E)$ can be expressed in terms of $\Sigma_{\lambda}\mathscr{G}_{\varepsilon}^{\pm}(\lambda,\lambda)$, and contains independent contributions from the ''plus'' and ''minus'' states. In this case only localized states emerge.

In an asymmetric double quantum well or when the impurity is not at the barrier's center, the off-diagonal elements of $\hat{\mathscr{V}}(p)$ are finite and the ''plus'' states mix with the ''minus'' states. In this case, depending on the relationship between Δ_T and the binding energy of the donor, two transformations of the bare ''plus'' or ''minus'' donor states are possible: transformation of a localized state to a resonant state if the former emerges against the background of the continuum, or discrete-level repulsion near the point of level crossing. Below we examine the contribution of these modifications of the spectrum to $\delta\rho_{\mathrm{imp}}(E)$ at E close to the energy of the ''plus'' donor ground state, when only the resonant contributions to the Green's function $g_{\varepsilon}^{+}(\mathbf{p},\mathbf{p}')$ should be taken into account. This approximation can be used when the tunnel coupling between the ''plus'' and ''minus'' states is weak; according to (23), such coupling can emerge when T/Δ_T is small (weak tunnel coupling between the l and r quantum wells), or when the difference $V_l(p)-V_r(p)$ is small (slightly asymmetric double quantum well). In the latter case, T/Δ_T may be of order unity, i.e., the results of this approximation can be used even when the interwell tunnel coupling is strong. Allowing only for the contribution of the ground state to the expansion (21), for the ''plus'' state we have

$$g_{\varepsilon}^{+}(\mathbf{p},\mathbf{p}')=\frac{\phi_0^{+}(\mathbf{p})\phi_0^{+}(\mathbf{p}')^{*}}{E_0^{+}-\varepsilon}, \tag{28}$$

where $\phi_0^{+}(\mathbf{p})$ and E_0^{+} are the eigenfunction and energy of the ground donor state, and the terms in the sum in (21) with $\lambda\neq 0$ have been discarded. As a result, from Eq. (26) for $\mathscr{G}_{\varepsilon}^{+}(\lambda,\lambda)$ it immediately follows that

$$\mathscr{G}_{\varepsilon}^{+}(0,0)=[E_0^{+}-W_{\varepsilon}^{-}(0,0)-\varepsilon]^{-1}. \tag{29}$$

The kernel W_{ε}^{+} in the integral equation (26) for $\mathscr{G}_{\varepsilon}^{-}$ is degenerate and is specified by

$$W_{\varepsilon}^{+}(\lambda,\lambda')=\frac{E(\lambda)E(\lambda')^{*}}{E_0^{+}-\varepsilon},$$

$$E(\lambda)=\sum_{\mathbf{p},\mathbf{p}'}\phi_{\lambda}^{-}(\mathbf{p})^{*}w(|\mathbf{p}-\mathbf{p}'|)\phi_0^{+}(\mathbf{p}'). \tag{30}$$

Using (3), we obtain a closed expression for $\mathscr{G}_{\varepsilon}^{-}(\lambda,\lambda')$:

$$\mathscr{G}_{\varepsilon}^{-}(\lambda,\lambda')=(E_{\lambda}^{-}-\varepsilon)^{-1}\left[\delta_{\lambda\lambda'}+\frac{E(\lambda)E(\lambda')^{*}}{E_{\lambda'}^{-}-\varepsilon}\times\left(E_0^{+}-\varepsilon-\sum_{\lambda_1}\frac{|E(\lambda_1)|^2}{E_{\lambda_1}^{-}-\varepsilon}\right)^{-1}\right]. \tag{31}$$

Thus, the donor contribution to the density of states can be expressed in terms of the sum

$$\sum_{\lambda}[\mathscr{G}_{\varepsilon}^{+}(\lambda,\lambda)+\mathscr{G}_{\varepsilon}^{-}(\lambda,\lambda)],$$

and to calculate this sum we require the variational solutions of (22) and the integrals in (29).

4.a. Resonant donor state

Here we examine the case where the donor state energy E_0^+, which can be found from Eq. (22) for the ''plus'' state, is greater than $-\Delta_T/2$, i.e., the level E_0^+ emerges against the background of the continuum formed by ''minus'' states. In this case ($w(p)\neq 0$), donor ''plus'' states mix with ''minus'' states of the subband, with the result that the discrete level is transformed into a resonant level. If this resonance is far from the edge of the continuum of the ''minus'' states, in calculating $W_\varepsilon^-(0,0)$ in (27) we can use the free Green's function

$$g_\varepsilon^-(\mathbf{p},\mathbf{p}')\simeq\delta_{\mathbf{pp}'}\left(\varepsilon_p-\frac{\Delta_T}{2}-\varepsilon\right)^{-1} \tag{32}$$

(i.e., the Green's function that ignores Coulomb corrections). Using $W_\varepsilon^-(0,0)$ from (27), we find that

$$W_\varepsilon^-(0,0)\simeq\sum_{\mathbf{p}}\frac{\Phi(\mathbf{p})}{\varepsilon_p-\Delta_T/2-\varepsilon},$$
$$\Phi(\mathbf{p})\simeq\left|\sum_{\mathbf{p}_1} w(|\mathbf{p}-\mathbf{p}_1|)\phi_0^+(\mathbf{p}_1)\right|^2. \tag{33}$$

The impurity contribution $\delta\rho_{\rm imp}(E)$ to the density of states for a narrow resonance is again given by (19). The shift of the peak energy E_R in relation to E_0^+ and the peak's half-width Γ are given by

$$E_R-E_0^+\simeq\mathscr{P}\sum_{\mathbf{p}}\frac{\Phi(\mathbf{p})}{\varepsilon_p-\Delta_T/2-E_0^+},$$
$$\Gamma\simeq\pi\sum_{\mathbf{p}}\Phi(\mathbf{p})\delta\left(\varepsilon_p-\frac{\Delta_T}{2}-E_0^+\right), \tag{34}$$

where $\mathscr{P}$ signifies the principal value of the integral.

We first calculate Γ with (34) for narrow quantum wells in which the width $\overline{d}$ of a double quantum well is much less than the Bohr radius a_B. Estimating the characteristic momentum p at $\hbar/a_B$, we expand the matrix elements up to first-order terms in $p\overline{d}/\hbar$, when $w(p)$ proves to be independent of p. If we now use the ground-state wave functions of the two-dimensional Coulomb problem to calculate $\Phi(\mathbf{p})$, we find that

$$\Gamma=R\left(\frac{\overline{d}}{a_B}\frac{T}{\Delta_T}\right)^2 F(z_D), \tag{35}$$

where $R=me^4/2\kappa^2\hbar^2$ is the effective Rydberg constant ($R\simeq 5.8$ meV for the parameters of GaAs). The function $F(z_D)$ for a double quantum well with flat bands, which is determined by the position of the impurity, is given by

$$F(z_D)=2\pi\left(\frac{d_l}{\overline{d}}\right)^2\left[\left(1+\frac{2z_D}{d_l}\right)^2-\frac{2(d+\overline{d})}{d_l}-\left(\frac{4}{\pi^2}\right)\cos\left(\frac{\pi z_D}{d_l}\right)^2\right]^2, \tag{36}$$

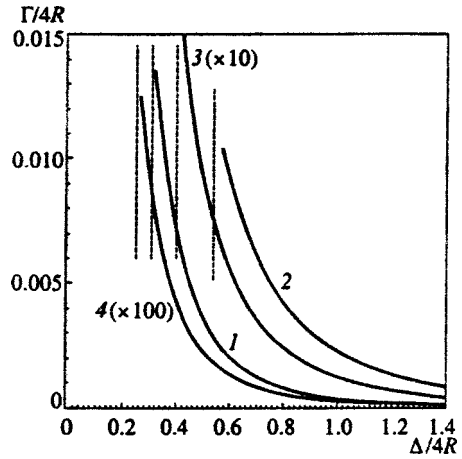

FIG. 4. Broadening of a resonant state as a function of level separation Δ for a 100/40/120-Å(AlGa)As double-quantum-well structure. The distance from the impurity to the left heteroboundary of the double quantum wells is 0 Å (curve *1*), 50 Å (curve *2*), 100 Å (curve *3*), and 120 Å (curve *4*). Energies are normalized to the binding energy $4R$ of the 2D Coulomb center. The vertical dashed lines indicate the energies at which a resonant state is transformed into a localized state.

where the impurity is assumed to be in the l quantum well, the position z_D is reckoned from the middle of the l well, and d is the barrier width. If the impurity is inside the barrier, (36) must be replaced by $F(z_D)=2\pi(8/\overline{d})^2[z_D+(d_l-d_r)/4]^2$, but if the impurity is outside the well, $F(z_D)=8\pi(1+d/\overline{d})^2$. In the adopted approximation, the principal contribution to the level shift is provided by large momenta, and we have the estimate

$$E_R-E_0^+\simeq\frac{\Gamma}{\pi}\ln\left|\frac{\varepsilon_m}{\Delta_T/2+E_0^+}\right|, \tag{37}$$

where $\varepsilon_m\simeq(\pi\hbar/\overline{d})^2/2m$ is the cutoff energy. Thus, in thin double quantum wells the shift of the narrow resonance is small.

The results of calculating the dependence of broadening on the level separation Δ for the case where the quantum-well width is comparable to the Bohr radius is depicted in Fig. 4 (the level separation Δ can be varied by applying a transverse electric field to the double quantum well). The energy E_0^+ was calculated by the variational method with a trial function $\phi_0^+(\rho)=\sqrt{8/\pi a_0^2}\exp(-2\rho/a_0)$, where a_0 is the variational parameter. The calculations were done for the AlGaAs/GaAs structure with quantum-well widths of 100 and 120 Å and a barrier width of 40 Å. For such a structure, in the absence of an external transverse electric field, $T/4R\simeq 0.05$ and $\Delta/4R\simeq 0.7$ ($4R$ is the binding energy of a two-dimensional Coulomb donor). Figure 4 shows that the dimensionless broadening $\Gamma/4R$ monotonically decreases as the level moves away from the edge of the continuum of the lower subband, which is shown by a vertical dotted line (near the edge of the continuum the broadening is large and the narrow-resonance approximation becomes invalid). Furthermore, broadening largely depends on the position z_D of the impurity: Γ increases severalfold if the impurity is shifted away from the outer boundary of the double quantum well

(curve *1*, $z_D=-d_l/2$) toward the center of the quantum well (curve *2*, $z_D=0$). A further shift of the impurity in the direction of the quantum-well–barrier interface (curve *3*, $z_D=d_l/2$) and toward the center of the barrier (curve *4*, $z_D=(d+d_l)/2$) leads to a rapid decrease in level broadening. At $z_D=(d+d_l)/2$ this broadening is due to the small difference in the quantum-well widths, with the result that a narrow peak emerges. This type of broadening is consistent with the results obtained for a point defect in Sec. 3.

4.b. Anticrossing of localized levels

We next consider a small level separation Δ_T for which the donor ground state E_0^+ turns out to lie below the bottom of the lowest subband, i.e., $E_0^*<-\Delta_T/2$. We examine the features of the energy spectrum that emerge when the level E_0^+ is close to the energy of the excited or ground states related to the lower subband. To calculate $\delta\rho_{\rm imp}(E)$ near the crossing of the ground states, in (27) we use the unperturbed Green's function $g_\varepsilon^-(\mathbf{p},\mathbf{p}')$ in the same approximations as in $g_\varepsilon^+(\mathbf{p},\mathbf{p}')$ in (28). This substitution yields $\mathscr{G}_\varepsilon^-(0,0)$ in a form similar to (29) with the kernel

$$W_\varepsilon^+(0,0)\simeq\frac{E^2(0)}{E_0^+-\varepsilon}. \tag{38}$$

Using these solutions for $\mathscr{G}_\varepsilon^\pm$, we obtain an expression for the impurity contribution to the density of states:

$$\delta\rho_{\rm imp}(E)\simeq\frac{2n_{\rm imp}}{\pi}\lim_{\varepsilon\to E+i0}\mathrm{Im}\,\frac{E_0^-+E_0^+-2\varepsilon}{(E_0^--\varepsilon)(E_0^+-\varepsilon)-E^2(0)}$$
$$=2n_{\rm imp}[\,\delta(E_--E)+\delta(E_+-E)], \tag{39}$$

where $E_\pm$ are the energy levels modified due to tunneling and defined as the poles of the fraction in (39). The energy E_- corresponds to the ground state and the energy E_+ to the first excited state of the donor in the double quantum well:

$$E_\pm=\frac{E_0^++E_0^-}{2}\pm\sqrt{\frac{(E_0^+-E_0^-)^2}{4}+E^2(0)}\,, \tag{40}$$

where $E_0^\pm$ are the eigenvalues determined by Eq. (22), and the energy $E(0)$ of level repulsion can be calculated using (30).

Let us study the dependence of $E_\pm$ on the level separation Δ (see Figs. 5 and 6) for a double quantum well with the same parameters as in Sec. 4a. In our calculations we use the same variational solutions as in calculating Γ. The functions $E_\pm(\Delta)$ vary significantly, depending on whether the impurity is in the interwell barrier or in a quantum well. If the impurity is inside the barrier, the $E_\pm$ vs. Δ curves demonstrate ordinary ''anticrossing'' (see the pairs of curves *4*, *4′* and *3*, *3′* in Fig. 5, where curves *4* and *3* correspond to the E_+ level and curves *4′* and *3′* to the E_- level). Figure 5 shows that the $E_\pm$ vs. Δ dependence differs only slightly from the behavior of the edge of the continuum, $\pm\Delta_T/2$ (cf. curves *3′* and *4′* and curve *0*, which corresponds to the edge of the spectrum, $-\Delta_T/2$). Similar nonmonotonic behavior of the binding energy of a donor and exciton was detected by Galbraith and Duggan[20] and Bayer and Timofeev.[21] Note

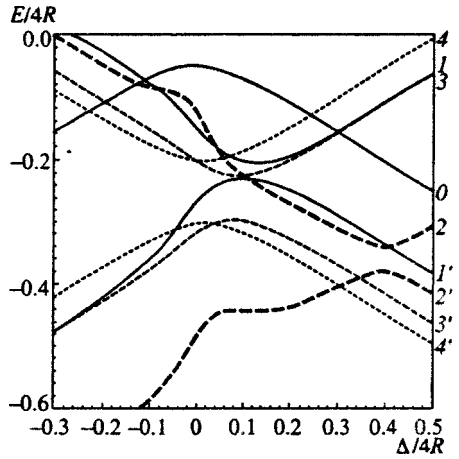

FIG. 5. Dependence of $E_\pm$ defined by (40) on Δ. For the same positions of the impurity as in Fig. 4, E_+ and E_- are given by the curves *1–4* and *1′–4′*, respectively. Curve *0* corresponds to the edge of the continuum, $-\Delta_T/2$.

that the segments of curves *1–4* above the edge of the continuum (curve *0*) correspond to the energy of the resonant states.

The $E_\pm$ vs. Δ dependence becomes more complicated when the impurity is inside a quantum well or at the outer heteroboundary (curves *2*, *2′* and *1*, *1′* in Fig. 5). The solutions $E_0^\pm$ of Eq. (22) obtained without accounting for the repulsion energy $E(0)$ are found to cross twice as Δ increases, as shown in Fig. 6 (the pairs of dashed and dotted curves in the lower panel). The $E_\pm$ vs. Δ dependence for an impurity localized at the center of the quantum well ($z_D=0$) and that for an impurity localized at a distance $3d_l/4$ from the outer heteroboundary ($z_D=d_l/4$) are depicted in Fig. 6 by solid curves. What is important here is that according to (23) and (30) the characteristic repulsion energy $E(0)$ in (40) resonantly increases as $\Delta\to0$ (this dependence is depicted in the upper panel of Fig. 6). Thus, for large values of Δ, where $E(0)$ varies monotonically, we have ordinary anticrossing, while for small values of Δ, where the crossing of the $E_0^\pm$ takes place simultaneously with a resonant in-

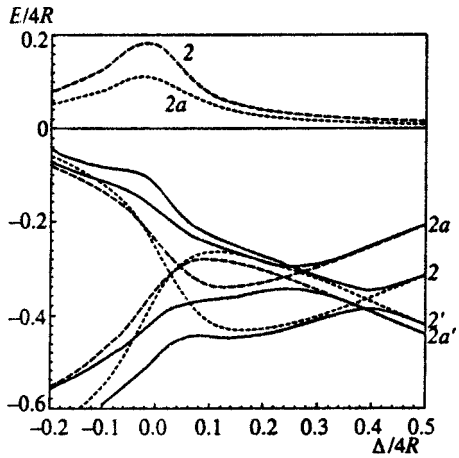

FIG. 6. Δ-dependence of $E_\pm$ for an impurity with $z_D=50$ Å (solid curves *2* and *2′*) and with $z_D=75$ Å (solid curves *2a* and *2a′*). The dashed and dotted curves in the lower panel correspond to energies $E_0^\pm$, and the curves in the upper panel represent the repulsion energies $E(0)$.

crease in $E(0)$, the repulsion of the levels $E_\pm$ increases, so that a singularity of the anticrossing type disappears.

Reasoning along similar lines, we can calculate the anticrossing of levels in the case where the ground-state energy E_0^+ is close to the energy E_l^- of the lowest excited states. Here $g_\varepsilon^+(\mathbf{p},\mathbf{p}')$ is given by (28) and the excited states are described only by the *s*- and *p*-contributionsto $g_\varepsilon^-(\mathbf{p},\mathbf{p}')$:

$$g_\varepsilon^-(\mathbf{p},\mathbf{p}')\simeq\sum_l \frac{\phi_{1l}^-(\mathbf{p})\phi_{1l}^-(\mathbf{p}')^*}{E_{1l}^- -\varepsilon}, \tag{41}$$

where $l=s,p$; E_{1p}^- is the energy of the twofold degenerate *p*-state; and E_{1s}^- is the energy of the excited *s*-state (note that in the approximation of a narrow quantum well $E_{1p}^-=E_{1s}^-$, i.e., we have a threefold degenerate excited state, as shown in Fig. 1b). Substituting these Green's functions into (29) and (31), we obtain

$$\mathscr{G}_\varepsilon^+(0,0)=\left[E_0^+-\varepsilon-\frac{|E(1s)|^2}{(E_{1s}^- -\varepsilon)}-\frac{2|E(1p)|^2}{(E_{1p}^- -\varepsilon)}\right]^{-1}, \tag{42}$$

$$\mathscr{G}_\varepsilon^-(1l,1l)=(E_{1l}^- -\varepsilon)^{-1}+\frac{|E(1l)|^2}{(E_{1l}^- -\varepsilon)^2}\mathscr{G}_\varepsilon^+(0,0).$$

Note that in this case two characteristic repulsion energies then emerge, $|E(1s)|$ and $|E(1p)|$, which are given by Eq. (30). Using (42) and transforming $\mathscr{G}_\varepsilon^+(0,0)+\Sigma_l\mathscr{G}_\varepsilon^-(1l,1l)$, we obtain

$$\delta\rho_{\rm imp}(E)\simeq\frac{2n_{\rm imp}}{\pi}\lim_{\varepsilon\to E+i0}{\rm Im}\left\{(E_{1p}^- -\varepsilon)^{-1}-\frac{d}{d\varepsilon}\ln[(E_1-\varepsilon)\times(E_2-\varepsilon)(E_3-\varepsilon)]\right\}$$

$$=2n_{\rm imp}\left[\delta(E_{1p}^- -E)+\sum_{j=1,2,3}\delta(E_j-E)\right], \tag{43}$$

where the E_j are the solutions of the equation

$$(E_{1p}^- -E)(\mathscr{E}_+ -E)(\mathscr{E}_- -E)-2|E(1p)|^2(E_{1s}^- -E)=0. \tag{44}$$

Here we have introduced the notation

$$\mathscr{E}_\pm=\frac{(E_0^+ +E_{1s}^-)}{2}\pm\sqrt{\frac{(E_0^+ -E_{1s}^-)^2}{4}+|E(1s)|^2}. \tag{45}$$

The value of the parameter E_{1s}^- lies between $\mathscr{E}_+$ and $\mathscr{E}_-$, i.e., the cubic equation (44) has three roots. Thus, we have found that anticrossing appears for both crossings of the levels, while there are no values of the parameters yielding one real and two complex-valued parameters. The study of a more complicated dependence of the solution on Δ (similar to those depicted in Fig. 6 for close-lying ground states) does not alter these conclusions, although the anticrossing pattern is more complicated. We do not list the results of our calculations here, since the observation of anticrossing of the excited and ground states is fairly complicated in doped double quantum wells due to the low splitting energies. This effect, however, might be of interest for low-density excitons.

5. CONCLUSION

In this paper we have discussed the special features of the impurity contribution to the density of states of double quantum wells, features that originate in the tunneling mixing of states of the *l* and *r* quantum wells. We found two qualitative changes in the impurity energy spectrum: the emergence of resonant states and the repulsion of the levels corresponding to the ground states or the ground and excited states. Earlier studies of donors dealt with the binding energy of the donor bound states as functions of the double-quantum-well parameters, with the energy found by standard variational calculations.[20,22,23] Below we discuss the possibility of these features showing up in experiments involving the optical and transport characteristics of the double quantum wells, and indicate the adopted approximations.

Note that the nonoverlapping of resonant states and the anticrossing of localized states discussed in this paper can be studied in experiments only for double-quantum-well doping levels that satisfy the condition $n_{\rm imp}\overline{r}^2\ll 1$ ($\overline{r}$ is the effective size of the donor, of the order of several Bohr radii). Due to the moderate electron concentrations, the sensitivity of submillimeter spectral measurements will be low. It would therefore be interesting to study the fundamental band-to-band transitions (by the photoluminescence or photoluminescence excitation spectra) in asymmetric double quantum wells. In such structures only the electronic states are tunnel-coupled, while the upper hole states are localized in one quantum well, since the separation of the hole extrema exceeds Δ_T.

Variations in Δ_T (controlled by a transverse electric field) can have a dramatic effect on the optical spectra of such structures due to the transformation of localized states into resonant states or, at a certain value of Δ_T, to the anticrossing effect. As noted in Sec. 1, excitonic peak broadening was detected by Fox *et al.*[13] and Oberli *et al.*,[14] but a detailed study of this effect has yet to be done. To our knowledge, the study of resonant states in Coulomb donors or structural defects (substitutional impurities or geometric faults in the heteroboundaries) has also yet to be conducted in double quantum wells. In such structures with nonideal heteroboundaries, the band-to-band transition edge broadens (this phenomenon was studied in Ref. 19), with the shape of the spectrum being highly dependent on the asymmetry of the scattering. For short-range defects, the shape of the optical spectrum also strongly depends on the defect localization in one or the other quantum well, since the hole states of only one quantum well participate in the transitions. The contribution of a narrow resonant state can also affect the longitudinal conductivity (or other transport coefficients) in selectively doped double quantum wells containing a δ-layer of doping impurities with a concentration of about $10^{11}\,{\rm cm}^{-2}$. Under a transverse voltage applied to the sample, the special features become evident when the energy of the resonant state coincides with the Fermi energy.

We now list the principal approximations adopted in the calculations. In describing impurity states we used the resonant tunneling approximation[19] and allowed for tunneling mixing of only the lowest pair of electronic levels in the

quantum wells, while all higher levels of the l and r quantum wells were ignored. The approximations of a parabolic energy spectrum and a homogeneous dielectric constant are common in structures of type I based on (GaAl)As or (GaIn)As; for such structures the models of Coulomb and short-range potentials used in this paper also hold. The one-center approximation (i.e., the overlap of wave functions at different centers is ignored) makes it possible to simplify calculations significantly if we write $\delta\rho_{\rm imp}(E)$ as a sum of separate impurity concentrations. The results are applicable only for small $n_{\rm imp}\overline{r^2}$ (see above). Here the states close to the edge of the continuum are ignored and the resonances are assumed narrow, so that the halfwidth Γ [see (19)] is small compared to the energy gap from the position of the resonance to the edge of the continuum. In solving the integral equation (24) we used the approximation in which tunneling modification of the spectrum (broadening of the resonance peak or the shift of levels due to mixing) is small compared to the characteristic energies in the absence of tunneling (binding energies of the impurities in separate quantum wells and level separation). These assumptions do not affect the qualitative picture of the donor states in double quantum wells.

Thus, we have studied the specific features of the impurity contribution to the density of states in double quantum wells for short-range defects and Coulomb donors. We have also discussed the feasibility of experimentally measuring the various effects, and consistency criteria for the results. We have noted that similar effects occur for excitonic states.

*)E-mail: zinovi@lab2.kiev.ua

[1] J. P. Eisenstein, Superlattices Microstruct. **12**, 107 (1992).
[2] T. Ohno, H. Sakaki, and M. Tsuchiya, Phys. Rev. B **49**, 11492 (1994).
[3] Y. Berk, A. Kamenev, A. Palevski, L. N. Pfeiffer, and K. W. West, Phys. Rev. B **51**, 2604 (1995).
[4] I. V. Butov, A. Zrenner, G. Abstreiter, A. V. Petinova, and K. Ebert, Phys. Rev. B **52**, 12153 (1995).
[5] Y. Huang and C. Lien, Phys. Low-Dim. Struct. **4/5**, 1 (1995).
[6] M. F. Krol, R. P. Leavitt, J. T. Pham, B. P. McGinnis, and N. Peyghambarian, Appl. Phys. Lett. **66**, 3045 (1995).
[7] V. B. Timofeev, A. V. Larionov, P. S. Dorozhkin, M. Bayer, A. Forchel, and J. Straka, JETP Lett. **65**, 877 (1997).
[8] A. M. Stoneham, *Theory of Defects in Solids: Electronic Structure and Defects in Insulators and Semiconductors*, Clarendon Press, Oxford (1975); J. Bourgoin and M. Lannoo, *Point Defects in Semiconductors II. Experimental Aspects*, Springer-Verlag, Berlin (1983).
[9] R. L. Greene and K. K. Bajaj, J. Vac. Sci. Technol. B **1**, 391 (1983).
[10] R. Ranganathan, B. D. McCombe, N. Nguyen, Y. Zhang, M. L. Rustgi, and W. J. Schaff, Phys. Rev. B **44**, 1423 (1991).
[11] A. B. Dzyubenko, Zh. Éksp. Teor. Fiz. **113**, 1446 (1998) [JETP **86**, 790 (1998)]; A. B. Dzyubenko and A. L. Yablonskiĭ, JETP Lett. **64**, 213 (1996).
[12] G. W. Bryant, Phys. Rev. B **47**, 1683 (1993).
[13] A. M. Fox, D. A. B. Miller, G. Livescu, J. E. Cunningham, and W. Y. Jan, Phys. Rev. B **44**, 6231 (1991).
[14] D. Y. Oberli, G. Bohm, G. Weinmann, and J. A. Brum, Phys. Rev. B **49**, 5757 (1994).
[15] S. Glutsch, F. Bechstedt, and D. S. Chemla, in *Proc. 23rd Int. Conf. on Phys. Semicond.*, M. Scheffler and R. Zimmermann (eds.), World Scientific, Singapore (1996).
[16] D. Birkedal, K. El. Sayed, G. Sanders, C. Spiegelberg, V. G. Lyssenko, C. Stanton, J. M. Hvam, V. B. Timofeev, and M. Bayer, Phys. Rev. B **54**, 10 316 (1996); G. Kh. Tartakovskiĭ, V. B. Timofeev, V. G. Lysenko, D. Birkedal, and J. Hvam, Zh. Éksp. Teor. Fiz. **112**, 1106 (1997) [JETP **85**, 601 (1997)].
[17] E. N. Economou, *Green's Functions in Quantum Physics*, Springer-Verlag, Berlin (1983).
[18] A. Yariv, C. Lindsey, and U. Sivan, J. Appl. Phys. **58**, 3669 (1985); F. T. Vas'ko and O. É. Raĭchev, Zh. Éksp. Teor. Fiz. **107**, 951 (1995) [JETP **80**, 539 (1995)].
[19] F. T. Vasko and O. E. Raichev, Phys. Rev. B **50**, 12 159 (1994); **51**, 7116 (1995).
[20] J. Galbraith and G. Duggan, Phys. Rev. B **40**, 5515 (1989).
[21] M. Bayer and V. B. Timofeev, Phys. Rev. B **54**, 8799 (1996).
[22] L. Shazhong and J. B. Khurgin, Phys. Rev. B **46**, 12 535 (1992).
[23] V. Takahashi, Y. Kato, S. Fukatsu, Y. Shivaki, and R. Ito, J. Appl. Phys. **76**, 2299 (1994).

Translated by Eugene Yankovsky

New method for detection of exciton Bose condensation using stimulated two-photon emission

Yu. E. Lozovik*) and A. V. Pushnov

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 18 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1353–1376 (April 1999)

An investigation is reported of stimulated two-photon emission by Bose-condensed excitons accompanied by a coherent two-exciton recombination, i.e., by simultaneous recombination of two excitons with opposite momenta leaving unchanged the occupation numbers of exciton states with momenta $\mathbf{p}\neq 0$. Raman light scattering (RLS) accompanied by a similar two-exciton recombination (or production of two excitons) is also analyzed. The processes under consideration can occur only if a system contains Bose condensate, so their detection can be used as a new method to reveal Bose condensation of excitons. The recoil momentum, which corresponds to a change in the momentum of the electromagnetic field in the processes, is transferred to phonons or impurities. If the recoil momentum is transmitted to optical phonons with frequency ω_0^s, whose occupation numbers are negligible, and the incident light frequency satisfies $\omega<2\Omega_-$, where $\Omega_-=\Omega-\omega_0^s$ is the difference frequency and Ω is the light frequency corresponding to the recombination of an exciton with zero momentum, stimulated two-photon emission and RLS with coherent two-exciton recombination give rise to a line at $2\Omega_- -\omega$ and an anti-Stokes component at $\omega+2\Omega_-$, respectively. For $\omega>2\Omega_-$ the RLS spectrum contains Stokes and anti-Stokes components at frequencies $\omega\pm 2\Omega_-$, whereas stimulated two-photon emission is impossible. Formulas for the cross sections at finite temperatures are obtained for the processes under consideration. Our estimates indicate that a spectral line at $2\Omega_- -\omega$, corresponding to the stimulated two-photon emission accompanied by coherent optical phonon-assisted two-exciton recombination can be experimentally detected in Cu_2O. © *1999 American Institute of Physics.* [S1063-7761(99)01404-3]

1. INTRODUCTION

The most interesting collective effects in systems of excitons are the anticipated exciton Bose condensation and superfluidity (see Refs. 1–7 and references therein). Recently a number of publications reported on the detection of Bose condensation and superfluidity of excitons in Cu_2O based on observations of changes in exciton luminescence spectra[8,9] and ballistic transport of excitons,[9–11] which have been discussed in the literature.[12–14] Observations of condensation of indirect excitons in coupled quantum wells under strong magnetic fields have also been reported (see Ref. 15, a theoretical discussion in Refs. 16–18, and references therein). In this connection, the detailed investigation of coherent exciton properties, whose detection could be used to reveal exciton Bose condensation, seems to be important.

If a system of excitons is in a Bose condensed state, the mean values of the annihilation (creation) operator of the exciton with zero momentum in the ground state are nonvanishing:

$$\langle N-1|Q_0|N\rangle=\langle N+1|Q_0^+|N\rangle=\sqrt{N_0}. \tag{1}$$

Here $|N\rangle$ is the ground state of the exciton system with the average number of excitons N, Q_0 is the annihilation operator of an exciton with zero momentum, and N_0 is the number of excitons in the condensate.

Equation (1) clearly shows that, as a result of the recombination (production) of an exciton with zero momentum, a system of Bose-condensed excitons goes over to the ground state, which differs from the initial one in the average number of excitons with momentum $\mathbf{p}=0$. The recombination of excitons with zero momentum is responsible for a peak (the so-called condensate peak) in the exciton luminescence spectrum at frequency $\Omega=[E_0(N)-E_0(N-1)]/\hbar$, where $E_0(N)$ is the energy of the ground state of the exciton system.

If the exciton–exciton interaction is nonvanishing, then, in addition to the mean values defined by Eq. (1), products of two annihilation (creation) operators of excitons with opposite momenta averaged over the ground state of the Bose-condensed exciton system (the so-called anomalous averages) are also nonvanishing:

$$\langle N-2|Q_{-p}Q_p|N\rangle\neq 0, \quad \langle N+2|Q_{-p}^+Q_p^+|N\rangle\neq 0. \tag{2}$$

In this paper we consider the unusual optical properties inherent in Bose-condensed state of interacting excitons due to the nonvanishing anomalous averages (2). It will be shown that coherent recombination or production, i.e., simultaneous annihilation or creation of two excitons with opposite momenta, corresponding to the anomalous averages (2) is possible due to interaction with the electromagnetic field. In such processes, the occupation numbers of excitons with $\mathbf{p}\neq 0$ are unchanged, and the final state of the excitons dif-

fers from the initial one only in the average number of excitons with zero momentum. In particular, after the two-exciton recombination, the average number of condensate excitons is reduced by two.

Coherent two-exciton recombination can contribute, for example, to the stimulated two-photon emission or to Raman light scattering (RLS) by Bose-condensed excitons. RLS can also be accompanied by coherent production of two excitons. In these processes, the momentum of the exciton–photon system is not conserved: the recoil momentum, equal to the change in the momentum of the electromagnetic field, is transferred to phonons or impurities.[19,20] In this paper we consider the processes in which the recoil momentum is transferred to two optical phonons. The prospects for such processes are probably best in the exciton system in Cu_2O crystal, which is one of most interesting crystals in view of the observation of exciton Bose condensation. In fact, the radiative recombination accompanied by the transmission of the recoil momentum to one optical phonon is typical for excitons in Cu_2O.[7] Using the energy and momentum conservation laws, one can prove that, in a defect-free crystal, coherent recombination of two excitons is possible only if the recoil momentum is transferred to two phonons.

At low temperatures, the occupation numbers of optical phonons are small, so it is most probable that the recoil momentum is transferred to two phonons produced in the process. If the phonon dispersion is negligible and the incident light frequency satisfies $\omega<2\Omega_-$, a line in the spectrum of the stimulated two-phonon emission at $2\Omega_- - \omega$ and an anti-Stokes component in the RLS spectrum at $\omega+2\Omega_-$ should appear. Here $\Omega_- = \Omega - \omega_0^s$ and ω_0^s is the optical phonon frequency. Both these lines correspond to coherent two-exciton recombination: the energy of the initial state of the system is higher than the energy of its final state by $2\hbar\Omega$, where Ω is the frequency corresponding to the recombination of an exciton with zero momentum. If $\omega>2\Omega_-$ holds, the RLS spectrum should contain the anti-Stokes component at $\omega+2\Omega_-$, which corresponds to coherent two-exciton recombination, and the Stokes component at $\omega-2\Omega_-$ due to coherent production of two excitons. Stimulated emission of two photons is impossible in this case. The lines at frequencies $|\omega\pm2\Omega_-|$ can appear only if the excitons are in the Bose-condensed state, and after a transition to the normal state these lines should disappear.

The paper is organized as follows. In Sec. 2 we consider stimulated two-photon emission with coherent two-exciton recombination accompanied by the transmission of the recoil momentum to phonons. The diagram technique is used to obtain the cross sections of two-photon processes involving coherent two-exciton recombination (or production) at finite temperatures. This approach allows one to express the appropriate elements of the *S*-matrix in a natural manner in terms of anomalous Green's functions of Bose-condensed excitons. The cross section of stimulated two-photon emission with coherent phonon-assisted two-exciton recombination is obtained, and its temperature dependence is studied. This dependence can be nonmonotonic under certain conditions. Namely, in a certain temperature range below T_c the cross section of stimulated two-photon emission can increase with the growth of temperature and can become even higher than it is at $T=0$. The causes of this unusual temperature dependence are investigated.

Section 3 is dedicated to RLS accompanied by coherent processes of two-exciton recombination or production. In Sec. 4 the possibility of experimental observation of the lines at frequencies $|\omega\pm2\Omega_-|$ corresponding to stimulated two-photon emission and RLS is analyzed. Our numerical estimates for excitons in Cu_2O indicate that a spectral line at $2\Omega_- - \omega$ corresponding to the stimulated optical phonon-assisted two-exciton recombination can be detected and, so can be used to reveal exciton Bose condensation.

2. STIMULATED TWO-PHOTON EMISSION ACCOMPANIED BY COHERENT TWO-EXCITON RECOMBINATION

The effective Hamiltonian describing phonon-assisted radiative recombination (production) of excitons can be expressed as follows (see Ref. 20 and Appendix A):

$$\begin{aligned}\hat{H}_L=\sum_{pq}\ [&L^{>}_{pq}e^{-i\Omega t}Q_p(t)c_q^+(t)b^+_{p-q}(t)\\ &+L^{<}_{pq}e^{-i\Omega t}Q_p(t)c_q^+(t)b_{q-p}(t)\\ &+L'^{>}_{pq}e^{-i\Omega t}Q_p(t)c_q(t)b^+_{p+q}(t)\\ &+L'^{<}_{pq}e^{-i\Omega t}Q_p(t)c_q(t)b_{-p-q}(t)+\text{H.c.}],\end{aligned}\tag{3}$$

where

$$L^{>(<)}_{pq}=i\sqrt{2\pi\omega_q}\mathbf{e}^*\cdot\mathbf{f}^{>(<)}_{pq},$$

$$L'^{>(<)}_{pq}=-i\sqrt{2\pi\omega_q}\mathbf{e}\cdot\mathbf{f}'^{>(<)}_{pq},$$

Ω is the frequency corresponding to the recombination of an exciton with zero momentum. The Hamiltonian (3) is written in the Heisenberg representation. Here $Q_p(t)=Q_p\times\exp[-i\epsilon(p)t]$ and $b_p(t)=b_p\exp(-i\omega_p^s t)$ are the annihilation operators of an exciton and a phonon with momentum p, respectively, and $c_q(t)=c_q\exp(-i\omega_q t)$ is the annihilation operator of a photon with momentum q (ω_q and $\mathbf{e}$ are the photon frequency and its polarization unit vector). The exciton energy is measured with respect to the bottom of the exciton band: $\epsilon(0)=0$. The effective matrix elements $\mathbf{f}^{>(<)}_{pq}$ and $\mathbf{f}'^{>(<)}_{pq}$ are responsible for the recombination of an exciton with momentum $\mathbf{p}$, which includes, in addition to the emission (absorption) of a photon with momentum $\mathbf{q}$, the simultaneous emission or absorption of a phonon.[1)] (see Ref. 20 and Appendix A).

By expanding the evolution operator

$$\hat{S}(t)=T_t\exp\left[-i\int_{-\infty}^{t}\hat{H}_L(t')dt'\right]$$

in powers of $\hat{H}_L$ and retaining terms up to second order, we obtain an expression for the elements of the *S*-matrix corresponding to phonon-assisted two-photon processes:

$$S_{n'n}=\frac{(-i)^2}{2!}\int\int_{-\infty}^{\infty}\langle n'|T_t\hat{H}_L(t')\hat{H}_L(t'')|n\rangle dt'dt'',\tag{4}$$

where n and n' label the initial and final states of the system composed of excitons and phonons + electromagnetic field.

Let us consider the two-photon emission by excitons in the Bose-condensed state due to coherent two-exciton recombination, i.e., a transition of the exciton system from state $|n\rangle_{\rm exc}=|n,N\rangle_{\rm exc}$ to the state $|m\rangle_{\rm exc}=|n,N-2\rangle_{\rm exc}$, which differs from the initial state in the average number of excitons with momentum $\mathbf{p}=0$. In this process, the change in the electromagnetic field momentum is $\mathbf{k}'+\mathbf{k}$, where $\mathbf{k}$ and $\mathbf{k}'$ are the momenta of emitted photons. The recoil momentum $\delta\mathbf{k}=-(\mathbf{k}'+\mathbf{k}')$ is entirely transferred to phonons, since the momentum of the exciton system is zero in both the initial and final states.

For the element of the S-matrix corresponding to coherent phonon-assisted two-exciton recombination, we have

$$
\begin{aligned}
(S_p)_{mn}=-\frac{1}{2}\int_{-\infty}^{\infty}\int dt'\,dt''\exp[-i\Omega(t'+t'')]\\
\times\{[L^{>}_{pk}L^{>}_{-pk'}\langle m|T_tQ_p(t')Q_{-p}(t'')|n\rangle_{\rm exc}\\
\times\langle f|T_tb^{+}_{p-k}(t')b^{+}_{-p-k'}(t'')|i\rangle_{\rm phon}\\
+L^{>}_{q-p,k}L^{>}_{p-q,k'}\langle m|T_tQ_{q-p}(t')Q_{p-q}(t'')|n\rangle_{\rm exc}\\
\times\langle f|T_tb^{+}_{-p-k'}(t')b^{+}_{p-k}(t'')|i\rangle_{\rm phon}]\\
\times\langle f|T_tc^{+}_{k}(t')c^{+}_{k'}(t'')|i\rangle_{\rm phot}\\
+[L^{>}_{-pk'}L^{>}_{pk}\langle m|T_tQ_{-p}(t')Q_{p}(t'')|n\rangle_{\rm exc}\\
\times\langle f|T_tb^{+}_{-p-k'}(t')b^{+}_{p-k}(t'')|i\rangle_{\rm phon}\\
+L^{>}_{p-q,k'}L^{>}_{q-p,k}\langle m|T_tQ_{p-q}(t')Q_{q-p}(t'')|n\rangle_{\rm exc}\\
\times\langle f|T_tb^{+}_{p-k}(t')b^{+}_{-p-k'}(t'')|i\rangle_{\rm phon}]\\
\times\langle f|T_tc^{+}_{k'}(t')c^{+}_{k}(t'')|i\rangle_{\rm phot}\},
\end{aligned}
\tag{5}
$$

where $\mathbf{q}=\mathbf{k}-\mathbf{k}'$. Here $|i\rangle_{\rm phot}=|0\rangle_{\rm phot}$ and $|f\rangle_{\rm phot}=|1_k,1_{k'}\rangle_{\rm phot}$ are the initial and final states of the electromagnetic field, respectively. Assuming that the phonons are optical and the lattice temperature $T_{\rm lat}$, which is, generally speaking, different from the exciton temperature T, is sufficiently small ($T_{\rm lat}\ll\omega_0^s$, where ω_0^s is the characteristic energy of optical phonons), we take $|i\rangle_{\rm phon}=|0\rangle_{\rm phon}$ and $|f\rangle=|1_{p-k},1_{-p-k'}\rangle_{\rm phon}$.

By averaging over the Gibbs distribution for the exciton system, we obtain the element of the S-matrix responsible for the two-photon emission that transforms the system from its state of thermodynamic equilibrium $|i\rangle_{\rm exc}=\Sigma_n\exp[(F-E_n(N)+\mu N)/T]|n,N\rangle_{\rm exc}$ to the state $|f\rangle_{\rm exc}=Q_0^2|i\rangle/N_0$:

$$
(S_p)_{fi}=\sum_n\exp[(F-E_n(N)+\mu N)/T](S_p)_{mn}. \tag{6}
$$

Expressing the S-matrix element (6) in terms of the anomalous Green's function of the excitons, we obtain

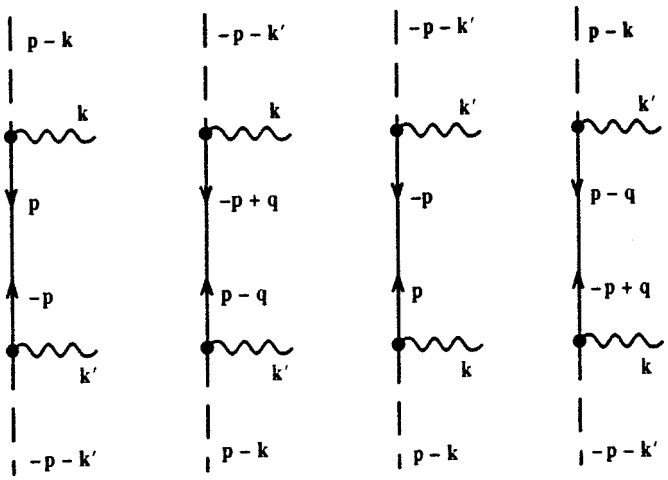

FIG. 1. Diagrams corresponding to two-photon emission accompanied by coherent phonon-assisted two-exciton recombination (the notation is explained in the text).

$$
\begin{aligned}
(S_p)_{fi}=-\frac{1}{2}\int_{-\infty}^{\infty}\int dt'\,dt''\exp[-i\Omega(t'+t'')]\\
\times\{[L^{>}_{pk}L^{>}_{-pk'}(n_0(T)\,\delta_p+i\hat{G}_{-p}(t'-t''))\\
\times\langle f|T_tb^{+}_{p-k}(t')b^{+}_{-p-k'}(t'')|i\rangle_{\rm phon}\\
+L^{>}_{q-p,k}L^{>}_{p-q,k'}(n_0(T)\,\delta_{p-q}+i\hat{G}_{p-q}(t'-t''))\\
\times\langle f|T_tb^{+}_{-p-k'}(t')b^{+}_{p-k}(t'')|i\rangle_{\rm phon}]\\
\times\langle f|T_tc^{+}_{k}(t')c^{+}_{k'}(t'')|i\rangle_{\rm phot}+L^{>}_{-pk'}L^{>}_{pk}(n_0(T)\,\delta_p\\
+i\hat{G}_p(t'-t''))\langle f|T_tb^{+}_{-p-k'}(t')b^{+}_{p-k}(t'')|i\rangle_{\rm phon}\\
+L^{>}_{p-q,k'}L^{>}_{q-p,k}(n_0(T)\,\delta_{p-q}+i\hat{G}_{q-p}(t'-t''))\\
\times\langle f|T_tb^{+}_{p-k}(t')b^{+}_{-p-k'}(t'')|i\rangle_{\rm phon}\langle f|T_tc^{+}_{k'}(t')\\
\times c^{+}_{k}(t'')|i\rangle_{\rm phot}\},
\end{aligned}
\tag{7}
$$

where $\delta_p=1$ at $p=0$ and $\delta_p=0$ for $p\neq0$. Here $\hat{G}_p(t'-t'')$ is the causal Green's function of Bose-condensed excitons at temperature T:

$$
\begin{aligned}
\hat{G}_p(t'-t'')=-i(1-\delta_p)\sum_n\exp[(F-E_n(N)+\mu N)/T]\\
\times\langle n,N-2|T_tQ_{-p}(t')Q_p(t'')|n,N\rangle_{\rm exc},
\end{aligned}
\tag{8}
$$

and the function $n_0(T)$ is the density of excitons in the condensate at this temperature.

The resulting element (7) of the S-matrix is expressed by the sum of diagrams shown in Fig. 1. The lines with oppositely directed arrows denote the causal anomalous Green's function of excitons in the Bose-condensed state for $T>0$ [if the momenta next to this line vanish, it corresponds to the function $n_0(T)$]. The wavy lines correspond to photon creation operators, and the dashed lines indicate phonon creation operators. The vertices on these diagrams correspond to

matrix elements $L^{>}_{pk}$, where p and k are the momenta of the exciton and photon lines originating at the vertex.[2)]

Integration over $t'-t''$ and t'' yields

$$(S_p)_{fi}=2\pi i T_{k'k}(\mathbf{p})[(\sqrt{2}-1)\delta(\mathbf{p}-\mathbf{q}/2)+1] \times\delta(\omega'+\omega+\omega^s_{p-k}+\omega^s_{-p-k'}-2\Omega), \tag{9}$$

where

$$T_{k'k}(\mathbf{p})=i\{L^{>}_{pk}L^{>}_{-pk'}[2\pi n_0(T)\delta_p\delta(\omega+\omega^s_{p-k}-\Omega) +i\hat{G}_p(\omega+\omega^s_{p-k}-\Omega)] +L^{>}_{q-p,k}L^{>}_{p-q,k'}[2\pi n_0(T)\delta_{p-q}\delta(\omega+\omega^s_{-p-k'} -\Omega)+i\hat{G}_{p-q}(\omega+\omega^s_{-p-k'}-\Omega)]\} \tag{10}$$

is the matrix element of the two-photon emission due to coherent phonon-assisted two-exciton recombination, which is similar to the scattering amplitude in the collision problem.[22] In deriving this equation, we have taken into account the fact that the anomalous Green's function is an even function of frequency and does not depend on the momentum direction. The sum in the brackets in Eq. (9) takes into account the fact that the momenta of emitted phonons are equal at $\mathbf{p}=\mathbf{q}/2$.

Let us limit our discussion to stimulated phonon-assisted two-photon emission with negligible phonon dispersion ($\omega^s_q=\omega^s_0$). It follows from Eq. (9) that stimulated two-photon emission of this kind gives rise to a line at frequency $2\Omega_- -\omega$, where $\Omega_-=\Omega-\omega^s_0$ and ω is the incident light frequency.[3)]

The differential cross section of stimulated two-photon emission corresponding to coherent phonon-assisted two-exciton recombination is given by

$$d\sigma^L=\frac{2\pi}{c}\left[\frac{1}{2}\sum_{\mathbf{p}\neq\mathbf{q}/2}|T_{k'k}(\mathbf{p})|^2 +2|T_{k'k}(\mathbf{q}/2)|^2\right]\frac{(2\Omega_- -\omega)^2}{(2\pi c)^3}do', \tag{11}$$

where

$$T_{k'k}(\mathbf{p})=i\{L^{>}_{pk}L^{>}_{-pk'}[2\pi n_0(T)\delta_p\delta(\omega-\Omega_-) +i\hat{G}_p(\omega-\Omega_-)]+L^{>}_{q-p,k}L^{>}_{p-q,k'}[2\pi n_0(T) \times\delta_{p-q}\delta(\omega-\Omega_-)+i\hat{G}_{p-q}(\omega-\Omega_-)]\}. \tag{12}$$

The factor 1/2 in front of the sum over $\mathbf{p}$ in Eq. (11) is introduced because the sum over all possible $\mathbf{p}$ includes the emission of two phonons with momenta $\mathbf{p}-\mathbf{k}$ and $-\mathbf{p}-\mathbf{k}'$ twice: $T_{k'k}(\mathbf{p})=T_{k'k}(-\mathbf{p}+\mathbf{q})$.

It is clear that for $\omega\neq\Omega_-$ the terms proportional to $n_0(T)$ do not contribute to the cross section (11). In this case, it is proportional to the anomalous Green's functions, which are determined, as is well known, not only by the presence of Bose condensate, but also by the interparticle interaction. Thus, stimulated two-photon emission corresponding to coherent phonon-assisted two-exciton recombination at $\omega\neq\Omega_-$ can take place only in a nonideal gas of Bose-condensed excitons.

Assuming that the condition $\omega\neq\Omega_-$ holds, we express cross section (11) as follows:

$$d\sigma^L=\frac{\omega(2\Omega_- -\omega)^3}{c^4}\left[\frac{1}{2}\sum_{p\neq q/2}|(s_p)_{nm}e'^*_n e^*_m|^2 +2|(s_{q/2})_{nm}e'^*_n e^*_m|^2\right]do', \tag{13}$$

where

$$(s_p)_{nm}=\hat{G}_p(\omega-\Omega_-)(f^{>}_{-pk'})_n(f^{>}_{pk})_m+\hat{G}_{p-q}(\omega-\Omega_-) \times(f^{>}_{-pk'})_n(f^{>}_{pk})_m \tag{14}$$

is the tensor of the two-photon emission corresponding to coherent phonon-assisted two-exciton recombination.

The causal Green's function $\hat{G}_p(\omega)$ is related to the advanced and retarded Green's functions by the following formula:[23]

$$\hat{G}_p(\omega)=\frac{1}{2}\left(1+\coth\frac{\omega}{2T}\right)\hat{G}^R_p(\omega) +\frac{1}{2}\left(1-\coth\frac{\omega}{2T}\right)\hat{G}^A_p(\omega). \tag{15}$$

By using Eq. (15) we arrive on the following

$$(s_p)_{nm}=\frac{1}{2}\left\{\left[\left(1+\coth\frac{\Delta\omega}{2T}\right)\hat{G}^R_p(\Delta\omega) +\left(1-\coth\frac{\Delta\omega}{2T}\right)\hat{G}^A_p(\Delta\omega)\right](f^{>}_{-pk'})_n(f^{>}_{pk})_m +\left[\left(1+\coth\frac{\Delta\omega}{2T}\right)\hat{G}^R_{p-q}(\Delta\omega) +\left(1-\coth\frac{\Delta\omega}{2T}\right)\hat{G}^A_{p-q}(\Delta\omega)\right] \times(f^{>}_{p-q,k'})_n(f^{>}_{q-p,k})_m\right\}, \tag{16}$$

where $\Delta\omega=\omega-\Omega_-$.

Using this expression, we calculate the sum over $\mathbf{p}$ in Eq. (13) for the cross section of stimulated two-photon emission:

$$\sum_{p\neq q/2}|(s_p)_{nm}e'^*_n e^*_m|^2 =\frac{1}{2}\sum_{p\neq q/2}\left\{2\left[\left(1+\coth^2\frac{\Delta\omega}{2T}\right)|\hat{G}^R_p(\Delta\omega)|^2 +\left(1-\coth^2\frac{\Delta\omega}{2T}\right)\mathrm{Re}[\hat{G}^R_p(\Delta\omega)]^2\right]|(\mathbf{e}'^*\cdot\mathbf{f}^{>}_{-pk'}) \times(\mathbf{e}^*\cdot\mathbf{f}^{>}_{pk})|^2+\left[\left(1+\coth\frac{\Delta\omega}{2T}\right)\hat{G}^R_p(\Delta\omega)+\left(1\right.\right.$$

$$-\coth\frac{\Delta\omega}{2T}\Big)\hat{G}^{R*}_{p-q}(\Delta\omega)\Big]\Big[\Big(1+\coth\frac{\Delta\omega}{2T}\Big)\hat{G}^{R*}_{p-q}(\Delta\omega)$$
$$+\Big(1-\coth\frac{\Delta\omega}{2T}\Big)\hat{G}^{R}_{p-q}(\Delta\omega)\Big](\mathbf{e}'^{*}\cdot\mathbf{f}^{>}_{-pk'})(\mathbf{e}^{*}\cdot\mathbf{f}^{>}_{pk})$$
$$\times(\mathbf{e}'\cdot\mathbf{f}^{>*}_{p-q,k'})(\mathbf{e}\cdot\mathbf{f}^{>*}_{q-p,k})\Big\}. \quad (17)$$

In deriving this formula, we have taken into account the relation between the advanced and retarded Green's functions on the real axis of ω: $G^{A}_{p}(\omega)=G^{R*}_{p}(\omega)$.

Further calculation of stimulated two-photon emission cross section (13) requires an expression for the retarded anomalous Green's function of excitons at a finite temperature. It can be obtained by analytically continuing the anomalous Green's function in the Matsubara representation to the upper half-plane of ω.

The anomalous Green's function of a Bose system in the Matsubara representation is given by the following expression:[24]

$$\hat{G}_p(\omega_s)=-\frac{(1-\delta_p)\Sigma^{02}_{\omega_s p}}{(i\omega_s-\epsilon_0(p)+\mu-\Sigma^{11}_{\omega_s p})(i\omega_s+\epsilon_0(p)-\mu+\Sigma^{11}_{-\omega_s,-p})+\Sigma^{20}_{\omega_s p}\Sigma^{02}_{\omega_s p}}, \quad (18)$$

where $\omega_s=2\pi sT$ and s is integer. Here $\epsilon_0(p)=p^2/2m$ and μ is the system chemical potential defined by the formula $\mu=[\Sigma^{11}_{\omega_s p}-\Sigma^{02}_{\omega_s p}]|_{\omega_s=p=0}$.

For $T\sim T_c$, where T_c is the Bose condensation temperature of an ideal Bose gas, the self-energy parts of a dilute Bose system with interparticle interaction can be expressed as follows:[24]

$$\Sigma^{11}_{\omega_s p}=\frac{8\pi}{m}na, \quad \Sigma^{20}_{\omega_s p}=\Sigma^{02}_{\omega_s p}=\frac{4\pi}{m}n_0(T)a, \quad (19)$$

where n is the total density of particles, a is the amplitude for their scattering by one another, and $n_0(T)$ is the total density of particles in the Bose condensate, which can be approximately calculated by the formula $n_0(T)=n[1-(T/T_c)^{3/2}]$.

Thus, the anomalous Green's function for a dilute exciton gas can be expressed as

$$\hat{G}_p(\omega_s)=(1-\delta_p)\frac{\zeta(T)}{\omega_s^2+\epsilon_p^2}, \quad (20)$$

where

$$\epsilon_p=\sqrt{\xi_p^2-\zeta^2(T)}, \quad \xi_p=\frac{p^2}{2m}+\zeta(T),$$
$$\zeta(T)=\mu(0)\Big[1-\Big(\frac{T}{T_c}\Big)^{3/2}\Big], \quad \mu(0)=\frac{4\pi na}{m},$$

m is the exciton mass. The parameter $\mu(0)$ is the chemical potential of the excitons at $T=0$.

Analytical continuation of $\hat{G}_p(\omega_s)$ to the upper half-plane yields an expression for the retarded anomalous Green's function:

$$\hat{G}^{R}_{p}(\omega)=-(1-\delta_p)\frac{\zeta(T)}{(\omega-\epsilon_p+i\Gamma_p/2)(\omega+\epsilon_p+i\Gamma_p/2)}. \quad (21)$$

Here $\Gamma_p=\tau_p^{-1}$ and τ_p is the lifetime of a quasiparticle with momentum p in the exciton system.

By substituting Eq. (21) in (17), one can easily find that the main contribution to the cross section of stimulated two-photon emission (13) at $|\Delta\omega|\gg\Gamma_p$ is due to the terms in which $\epsilon_p\sim|\Delta\omega|$. Therefore, matrix elements $\mathbf{f}^{>}_{-pk'}$ and $\mathbf{f}^{>}_{pk}$ can be replaced by their values corresponding to the momentum p_L that satisfies the condition $\epsilon(p_L)=\Delta\omega$ and carried out of the integrand. Moreover, if the condition $p_L\gg q$ is fulfilled, one can set $q=0$ in the sum over p in Eq. (17). Thus, we have

$$\sum_{\mathbf{p}}|(s_p)_{nm}e'^{*}_{n}e^{*}_{m}|^2=2\Big[\Big(1+\coth^2\frac{\Delta\omega}{2T}\Big)\sum_{p}|\hat{G}^{R}_{p}(\Delta\omega)|^2$$
$$+\Big(1-\coth^2\frac{\Delta\omega}{2T}\Big)\sum_{p}\mathrm{Re}[\hat{G}^{R}_{p}(\Delta\omega)]^2\Big]$$
$$\times|f_n(\omega'_L)f_m(\omega_L)e'^{*}_{n}e^{*}_{m}|^2, \quad (22)$$

where

$$\mathbf{f}(\omega_L)=\frac{1}{4\pi}\int\mathbf{f}^{>}(p_L,k)do_{p_L},$$
$$\mathbf{f}(\omega'_L)=\frac{1}{4\pi}\int\mathbf{f}^{>}(p_L,k')do_{p_L}$$

are the matrix elements averaged over the directions of vector $\mathbf{p}_L$.

Replacing the summation in Eq. (22) by the integration over $\mathbf{p}$, we obtain

$$\sum_{\mathbf{p}}|\hat{G}^{R}_{p}(\Delta\omega)|^2=\int_0^\infty\frac{d^3p}{(2\pi)^3}\frac{\zeta^2(T)}{|(\Delta\omega+i\Gamma_p/2)^2-\epsilon_p^2|^2}. \quad (23)$$

This (23) integral diverges as $\Gamma_p\to0$. Representing it as a sum of two integrals each of which converges at $\Gamma_p\to0$ and replacing the integration over $\mathbf{p}$ by the integration over $t=\xi_p/\zeta(T)$ yields

$$\sum_{\mathbf{p}} |\hat{G}_p^R(\Delta\omega)|^2 = \frac{\sqrt{2m^3/\zeta(T)}}{2\pi^2(\beta_+^2-\beta_-^2)} \times\left[\int_1^\infty \frac{dt\sqrt{t-1}}{t^2-\beta_+^2} - \int_1^\infty \frac{dt\sqrt{t-1}}{t^2-\beta_-^2}\right], \tag{24}$$

where

$$\beta_\pm^2 = (\alpha_L \pm i\gamma_L)^2+1, \quad \alpha_L = \frac{|\Delta\omega|}{\zeta(T)},$$

$$\gamma_L = \frac{\Gamma_L}{2\zeta(T)}, \quad \Gamma_L = \Gamma_{p_L}.$$

Thus, in calculating the integrals on the right-hand side of the resulting equation, one can set $\beta_\pm^2 = \beta^2 \pm i\delta$. As a result, we obtain

$$\sum_{\mathbf{p}} |\hat{G}_p^R(\Delta\omega)|^2 = \frac{\sqrt{2m^3\zeta(T)(\sqrt{\alpha_L^2+1}-1)}}{4\pi\alpha_L\Gamma_L\sqrt{\alpha_L^2+1}}. \tag{25}$$

The second sum over $\mathbf{p}$ in Eq. (22) converges even as $\Gamma_p \to 0$. Therefore, if $|\Delta\omega| \gg \Gamma_p$ holds we can set $\Gamma_p = 0+$ in this sum. In this case, we have

$$\mathrm{Re}\sum_p [\hat{G}_p^R(\Delta\omega)]^2 = -\frac{\sqrt{2m^3/\zeta(T)}}{16\pi} \times\frac{\sqrt{\sqrt{\alpha_L^2+1}-1}(\sqrt{\alpha_L^2+1}+2)}{\alpha_L\sqrt{(\alpha_L^2+1)^3}}. \tag{26}$$

It is clear that for $|\Delta\omega| \gg \Gamma_p$ the following relation takes place:

$$\left|\sum_{\mathbf{p}} \mathrm{Re}[G_p^R(\Delta\omega)]^2\right| \ll \sum_p |G_p^R(\Delta\omega)|^2.$$

Thus,

$$\sum_{\mathbf{p}} |(s_p)_{nm} e'^*_n e^*_m|^2 = \frac{\sqrt{2m^3\zeta(T)(\sqrt{\alpha_L^2+1}-1)}}{2\pi\alpha_L\Gamma_L\sqrt{\alpha_L^2+1}} \times\left(1+\coth^2\frac{\Delta\omega}{2T}\right) \times|f_n(\omega_L')f_m(\omega_L)e'^*_m e^*_n|^2. \tag{27}$$

By substituting this expression in the formula for the differential cross section (13), we obtain

$$d\sigma^L = \frac{\omega(2\Omega_- - \omega)^3}{4\pi c^4} \frac{\sqrt{2m^3\zeta(T)(\sqrt{\alpha_L^2+1}-1)}}{\alpha_L\Gamma_L\sqrt{\alpha_L^2+1}} \times\left(1+\coth^2\frac{\Delta\omega}{2T}\right)|f_n(\omega_L')f_m(\omega_L)e'^*_n e^*_m|^2 do'. \tag{28}$$

If the exciton–phonon system is isotropic and the incident light is monochromatic and linearly polarized, one has $|e_m^* f_m(\omega_L)|^2 = \mathbf{f}^2(\omega_L)/3$. Summing over the polarizations of photon ω' and integrating over the directions of its momentum (note that in the case of stimulated two-photon emission the photon ω is identical to the incident one), we obtain the total cross section of stimulated two-photon emission corresponding to the coherent phonon-assisted two-exciton recombination:

$$\sigma^L(\omega,T) = \frac{\omega(2\Omega_- - \omega)^3}{c^4} \frac{\sqrt{8m^3\zeta(T)(\sqrt{\alpha_L^2+1}-1)}}{9\alpha_L\Gamma_L\sqrt{\alpha_L^2+1}} \times\left(1+\coth^2\frac{\Delta\omega}{2T}\right)\mathbf{f}^2(\omega_L)\mathbf{f}^2(\omega_L'). \tag{29}$$

Note that, if the conditions $\Delta\omega \ll \Omega_-$, $\mu(0) \ll \Omega$, and $\tau^L = \mathrm{const}$ are fulfilled, then at a given ratio between the exciton chemical potential $\mu(0)$ at zero temperature and twice the temperature of their Bose condensation, $\gamma = \mu(0)/2T_c$ the parameter $\sigma^L(\Delta\omega,T)/\sigma^L(0,0)$ is uniquely determined by two quantities, $x = \Delta\omega/2T_c$ and $y = T/T_c$:

$$\frac{\sigma^L(\Delta\omega,T)}{\sigma^L(0,0)} = \frac{z^2\sqrt{\sqrt{x^2+z^2}-z}}{|x|\sqrt{2\gamma(x^2+z^2)}}\left(1+\coth^2\frac{x}{y}\right), \tag{30}$$

where $z = \gamma(1-y^{3/2})$.

The dependence of the cross section (29) on frequency (strictly speaking, on the difference between the incident light frequency ω and Ω_-) is shown in Fig. 2a for different temperatures of the exciton subsystem. This cross section as a function of temperature at different fixed values of the difference $\Delta\omega = \omega - \Omega_-$ is shown in Fig. 2b. All the curves in Fig. 2 correspond to $\gamma = 0.3$, and it is assumed that $\tau^L = \mathrm{const}$. It is clear that for $|\Delta\omega| \ll T_c$ and $T < T_c$ there is a temperature range where the cross section (29) of stimulated two-photon emission is a nonmonotonic function of temperature: σ^L increases with the growth of temperature and can even become larger than it is at $T = 0$.

The reason for this unusual temperature dependence is the following. The cross section (29) of stimulated two-photon emission is determined by two quantities that depend on the temperature differently, namely, through $\zeta(T)$, which is proportional to the number of excitons in the condensate, and through the occupation numbers of quasiparticle levels of the exciton system with the quasiparticle energy $\epsilon(p_L) = |\Delta\omega|$. Specifically, the density of the condensate and hence $\zeta(T)$ decrease as the temperature increases. This, in turn, reduces the cross section (29). On the other hand, using Bogoliubov's u–v transforms, one can easily show that coherent two-exciton recombination, which is a second-order process with respect to the Hamiltonian (3), proceeds via intermediate states of the exciton system containing one particle more (less) than the state of thermodynamic equilibrium (see also Refs. 19 and 20). The cross section of stimulated two-photon emission corresponding to coherent two-exciton recombination is proportional to

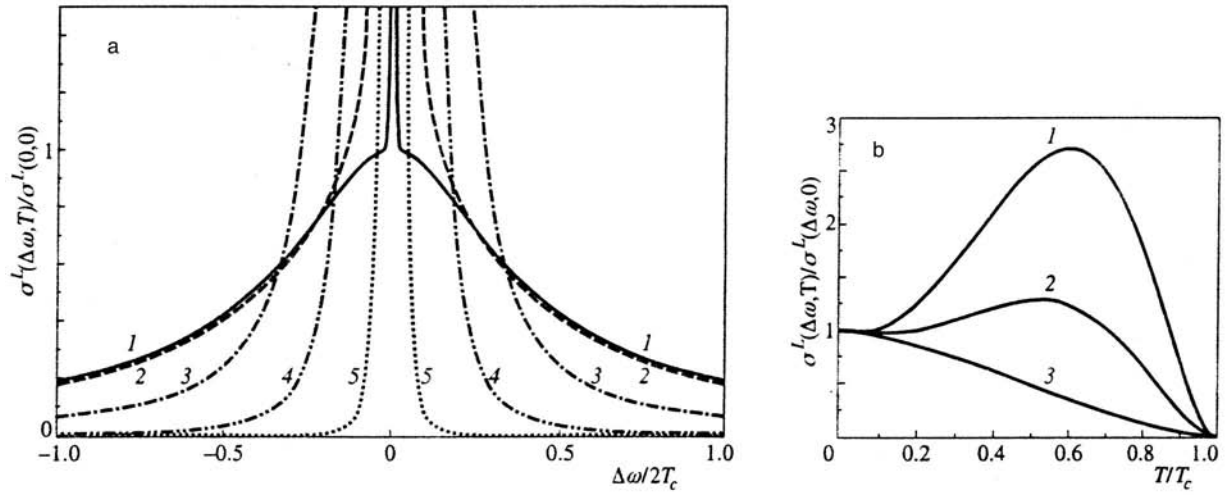

FIG. 2. Cross section of stimulated two-photon emission accompanied by coherent optical phonon-assisted two-exciton recombination: (a) as a function of the difference $\Delta\omega=\omega-\Omega_-$ between the incident light frequency ω and Ω_- at various temperatures T of the exciton system: (*1*) $T/T_c=0.01$; (*2*) 0.10; (*3*) 0.60; (*4*) 0.90; (*5*) 0.99; (b) as a function of the exciton system temperature T at various $\Delta\omega$: (*1*) $|\Delta\omega|/2T_c=0.2$; (*2*) 0.3; (*3*) 0.9. The curves were plotted using Eq. (30). For all curves $\mu(0)/2T_c=0.3$, and τ^L is assumed to be constant.

$$(n_{p_L}+1)^2+n_{p_L}^2=\frac{1}{2}\left(1+\coth^2\frac{\Delta\omega}{2T}\right),$$

where $n_{p_L}=[\exp(\epsilon_{p_L}/T)-1]^{-1}$ is the occupation number of the quasiparticle state with energy $\epsilon(p_L)=|\Delta\omega|$ in the exciton system. As the temperature increases, n_p also rises, which increases the cross section (29). If this tendency dominates, the cross section of stimulated two-photon emission corresponding to coherent two-exciton recombination should increase with the temperature. In fact, the tendency to decrease the cross section should dominate sooner or later as $T\to T_c$, since it must vanish at $T=T_c$.

Note that the temperature dependence of the cross section (29) of the stimulated two-photon emission accompanied by coherent two-exciton recombination has been calculated in the approximation (19), which is correct only in a narrow temperature interval about the Bose condensation temperature T_c, which is assumed to equal the Bose condensation temperature in an ideal Bose gas. Although this approximation allows one to reproduce formally our results[20] for $T=0$, in the intermediate temperature range the curve of the cross section $\sigma^L(\Delta\omega,T)$ versus temperature should be different from that plotted in Fig. 2. Nonetheless, the conclusion about the nonmonotonic temperature dependence of the cross section of stimulated two-photon emission is valid. For example, at $\Delta\omega/2T_c=0.2$ we have $\sigma^L(\Delta\omega,T)>\sigma^L(\Delta\omega,0)$ even for $T_c-T\ll T_c$ (Fig. 2b), where the approximation (19) is correct.

3. RAMAN LIGHT SCATTERING

Coherent two-exciton recombination can accompany not only in the stimulated two-photon emission but also the Raman light scattering (RLS). Abrikosov and Falkovsky[25] analyzed RLS in a superconductor, whose analogue in a semiconductor was RLS by a dense electron–hole plasma with coupling between electrons and holes (a phase transition in this system was studied by Keldysh and Kopaev;[26] see also the review by Kopaev[27] and references therein). But we are discussing the case of a low density of electrons and holes (exciton gas). Moreover, it is essential for the case of RLS under consideration that the electron–hole system not be in equilibrium, because this is the situation when coherent two-exciton recombination (production) can cause the exciton system to undergo a transition to a state with a lower (higher) energy. In the case of such RLS with the transfer of the recoil momentum to the two optical phonons energy conservation is described by the formula

$$\omega+2\Omega_-=\omega'. \tag{31}$$

The case considered here corresponds to the appearance of an anti-Stokes component at frequency ω' defined by this formula.

In addition, an RLS process with coherent two-exciton production is also possible, and energy conservation in this case is described by the equation[4)]

$$\omega-2\Omega_-=\omega'. \tag{32}$$

This formula determines the frequency ω' of the Stokes component corresponding to this Raman scattering. It is clear that RLS with coherent phonon-assisted two-exciton production is possible only for $\omega>2\Omega_-$. Stimulated two-photon emission corresponding to the coherent phonon-assisted two-exciton recombination is impossible in this case.

The analysis of RLS accompanied by coherent two-exciton recombination (or production) is similar to that of stimulated two-photon emission with coherent two-exciton recombination. Since the formulas for the cross section of RLS with coherent two-exciton recombination or production are lengthy, here we only indicate how these formulas can be derived from Eq. (29) using appropriate substitutions.

1. The cross section of RLS accompanied by coherent phonon-assisted two-exciton recombination is obtained by replacing some variables in Eq. (29):

$$\mathbf{f}(\omega_L)\to\mathbf{f}'(\tilde{\omega}_L),\quad \mathbf{f}(\omega_L')\to\mathbf{f}(\tilde{\omega}_L'),\quad \omega\to-\omega,$$

$$\Delta\omega\to\omega+\Omega_-,\quad \alpha_L\to\tilde{\alpha}_L,\quad \Gamma_L\to\tilde{\Gamma}_L.$$

Here $\tilde{\alpha}_L=(\Omega_-+\omega)/\zeta(T)$, $\tilde{\Gamma}_L$ is the reciprocal lifetime of a quasiparticle with energy $\epsilon(\tilde{p}_L)=\Omega_-+\omega$ in the exciton system,

$$\mathbf{f}'(\tilde{\omega}_L)=\frac{1}{4\pi}\int \mathbf{f}'^{>}(\tilde{p}_L,k)do_{\tilde{p}_L},$$

$$\mathbf{f}(\tilde{\omega}_L')=\frac{1}{4\pi}\int \mathbf{f}^{>}(\tilde{p}_L,k')do_{\tilde{p}_L}.$$

2. The cross section of RLS due to coherent phonon-assisted two-exciton production ($\omega>2\Omega_-$) is derived from Eq. (29) by substituting

$$\mathbf{f}(\omega_L)\to\mathbf{f}'(\omega_L),\quad 2\Omega_--\omega\to\omega-2\Omega_-,$$

where

$$\mathbf{f}'(\omega_L)=\frac{1}{4\pi}\int \mathbf{f}'^{>}(p_L,k)do_{p_L}.$$

4. POSSIBILITY OF EXPERIMENTAL DETECTION OF TWO-PHOTON PROCESSES ACCOMPANIED BY COHERENT TWO-EXCITON RECOMBINATION

Let us analyze the possibility of experimentally detecting stimulated two-photon emission and RLS accompanied by coherent phonon-assisted two-exciton recombination. First we consider stimulated two-photon emission.

The light intensity $I^L(\omega')$ at frequency $\omega'=2\Omega_--\omega$ resulting from stimulated two-photon emission with transfer of the recoil momentum to the optical phonons is given by the expression

$$I^L(\omega')=\frac{\omega'}{\omega}\sigma^L(\omega)I(\omega), \tag{33}$$

where $\sigma^L(\omega)$ is the cross section of this process [Eq. (29)] and $I(\omega)$ is the intensity (in W/cm^2) of incident light of frequency ω.

The intensity (33) can be expressed as a sum of two terms:

$$I^L(\omega')=\Delta I^L(\omega')+\tilde{I}^L(\omega'), \tag{34}$$

where $\tilde{I}^L(\omega')$ is the intensity of the two-photon emission resulting from two consecutive processes: the spontaneous emission at frequency $\omega'=2\Omega_--\omega$ and the subsequent stimulated emission at frequency ω, each of which satisfies the energy conservation law.

If the incident light frequency satisfies $\omega>\Omega_-$, then $\omega'<\Omega_-$ holds. In this case, the spontaneous emission at frequency $\omega'=2\Omega_--\omega<\Omega_-$ is due to exciton recombination with production of a Bogoliubov quasiparticle with momentum $\mathbf{p}_L$ that satisfies the condition $\epsilon(p_L)=\Delta\omega$ (see Appendix B, $\Delta\omega=-\Delta\omega'$). The spontaneous recombination of excitons produces in the exciton system $I_s^L(\omega')/\omega'$ quasiparticles with energy $\epsilon(p_L)=\Delta\omega$ per unit time, where $I_s^L(\omega')$ is the luminescence intensity (57) (see Appendix B and Ref. 28). These quasiparticles disappear in a time of order of τ^L. The disappearance of some of these quasiparticles is accompanied by the stimulated recombination of excitons and the induced emission of light at frequency ω. Thus, for $\omega>\Omega_-$ the intensity $\tilde{I}^L(\omega')$ is given by the relation

$$\tilde{I}^L(\omega')=\frac{\tau^L}{\tau_r^L}I_s^L(\omega'), \tag{35}$$

where τ_r^L is the lifetime of the quasiparticle with energy $\epsilon(p_L)$ with respect to its recombination accompanied by stimulated emission at frequency ω, provided that the exciton system contains one quasiparticle with momentum $\mathbf{p}_L$ more than it does in the state of thermodynamic equilibrium. The time τ_r^L can be easily calculated using Fermi's ''golden rule'':

$$\frac{1}{\tau_r^L}=\frac{(2\pi)^2}{3c}\mathbf{f}^2(\omega_L)u_{p_L}^2(n_{p_L}+1)I(\omega),$$

$$u_{p_L}^2=\frac{1}{2}\left(\frac{\sqrt{\alpha_L^2+1}}{\alpha_L}+1\right),\quad n_{p_L}=\frac{1}{e^{\Delta\omega/T}-1}, \tag{36}$$

where u_{p_L} is Bogoliubov's coefficient and n_{p_L} is the distribution function of quasiparticles with energy $\epsilon(p_L)=\Delta\omega$ at temperature T.

If the incident light frequency satisfies $\omega<\Omega_-$ and hence $\omega'>\Omega_-$, the situation is similar to that discussed above. In this case, the spontaneous emission at frequency ω' is due to the recombination of an exciton accompanied by the disappearance of one Bogoliubov quasiparticle with energy $\epsilon(p_L)=-\Delta\omega$ in the exciton system. For $\omega'>\Omega_-$, the number of quasiparticles of energy $\epsilon(p_L)=-\Delta\omega$ which disappear per unit time as a result of spontaneous recombination of excitons is $I_s^L(\omega')/\omega'$. In a time of order of τ^L, the missing quasiparticles are replaced by new ones, some of which are accompanied by stimulated emission at frequency ω. Thus, for $\omega<\Omega_-$ we have

$$\tilde{I}^L(\omega')=\frac{\tau^L}{\tau_c^L}I_s^L(\omega'), \tag{37}$$

where τ_c^L is the lifetime of an exciton with momentum $\mathbf{p}_L$ with respect to stimulated recombination, which results in both stimulated emission at frequency ω and production of a quasiparticle of energy $\epsilon(p_L)=-\Delta\omega$, provided that the exciton system contains one quasiparticle with momentum $\mathbf{p}_L$ less than it does in the state of thermodynamic equilibrium. Using Fermi's ''golden rule,'' we obtain the following expression for τ_c^L:

$$\frac{1}{\tau_c^L}=\frac{(2\pi)^2}{3c}\mathbf{f}^2(\omega_L)v_{p_L}^2n_{p_L}I(\omega),$$

$$v_{p_L}^2=\frac{1}{2}\left(\frac{\sqrt{\alpha_L^2+1}}{\alpha_L}-1\right). \tag{38}$$

Using Eqs. (35)–(38) and (57) from Appendix B, we obtain the intensity $\tilde{I}^L(\omega')$ in the general case:

$$\tilde{I}^L(2\Omega_- - \omega) = \frac{2\Omega_- - \omega}{\omega}\tilde{\sigma}^L(\omega)I(\omega),$$

$$\tilde{\sigma}^L(\omega) = \tau^L \frac{\omega(2\Omega_- - \omega)^3}{c^4} \frac{\sqrt{2m^3\zeta(T)(\sqrt{\alpha_L^2+1}-1)}}{18\alpha_L\sqrt{\alpha_L^2+1}} \times \mathbf{f}^2(\omega_L)\mathbf{f}^2(\omega_L')\left[\mathrm{sign}(\Delta\omega) + \coth\frac{|\Delta\omega|}{2T}\right]^2. \tag{39}$$

The spectral line at frequency $\omega' = 2\Omega_- - \omega$ due to the stimulated two-photon emission accompanied by coherent phonon-assisted two-exciton recombination will be superposed on the luminescence spectrum of Bose-condensed excitons. It clearly follows from Eqs. (35) and (37) that $\tilde{I}^L(\omega')$ determines a fraction of the spontaneous emission intensity $I_s^L(\omega')$. Thus, the total intensity of the emission at frequency ω' can be expressed as follows:

$$I_{\mathrm{tot}}^L(\omega') = \Delta I^L(\omega') + I_s^L(\omega'), \tag{40}$$

where $I_s^L(\omega')$ is the luminescence intensity at frequency ω' in the absence of incident light of frequency ω, $\Delta I^L(\omega')$ is the observed light intensity at frequency ω' due to stimulated two-photon emission with coherent two-exciton recombination. By substituting the cross section (29) in Eq. (33) and using Eqs. (34) and (39), we obtain the observed light intensity $\Delta I^L(\omega')$:

$$\Delta I^L(2\Omega_- - \omega) = \frac{2\Omega_- - \omega}{\omega}\Delta\sigma^L(\omega)I(\omega),$$

$$\Delta\sigma^L(\omega) = \tau^L \frac{\omega(2\Omega_- - \omega)^3}{c^4} \frac{\sqrt{8m^3\zeta(T)(\sqrt{\alpha_L^2+1}-1)}}{9\alpha_L\sqrt{\alpha_L^2+1}} \times \mathbf{f}^2(\omega_L)\mathbf{f}^2(\omega_L')\left[1 + \coth^2\frac{\Delta\omega}{2T} - \frac{1}{4} \times \left(\mathrm{sign}(\Delta\omega) + \coth\frac{|\Delta\omega|}{2T}\right)^2\right]. \tag{41}$$

One can easily prove that $1/2 \leqslant \Delta\sigma^L(\omega)/\sigma^L(\omega) \leqslant 1$. In particular, at $T=0$ we have $\Delta\sigma^L(\omega) = \sigma^L(\omega)$ at $\omega < \Omega_-$ and $\Delta\sigma^L(\omega) = \sigma^L(\omega)/2$ at $\omega > \Omega_-$.

Using Eq. (29), we can estimate the cross section σ^L of stimulated two-photon emission. In CGS units this expression has the form

$$\sigma^L(\omega,T) = \tau^L V \frac{\omega(2\Omega_- - \omega)^3}{9c^4\hbar^4} \frac{\sqrt{8m^3\zeta(T)(\sqrt{\alpha_L^2+1}-1)}}{\alpha_L\sqrt{\alpha_L^2+1}} \times \left(1 + \coth^2\frac{\Delta\omega}{2T}\right)\mathbf{f}^2(\omega_L)\mathbf{f}^2(\omega_L'), \tag{42}$$

where V is the volume of excitons interacting with incident light and $\alpha_L = \hbar|\Delta\omega|/\zeta(T)$.

We shall consider as an example a system of Bose-condensed excitons in Cu_2O at zero temperature. The exciton effective mass in this crystal is $m = 2.7m_e$, the characteristic exciton radius is $a = 7$ Å, and the photon energy corresponding to the recombination of an exciton with zero momentum is $\hbar\Omega \simeq 2$ eV. The optical recombination of an exciton in Cu_2O is typically assisted by production of an optical phonon of energy $\hbar\omega_0^s \simeq 10$ meV with a negligible dispersion.

Let us estimate the exciton chemical potential at $T=0$ by the formula

$$\mu(0) = \frac{4\pi\hbar^2}{m}na,$$

where n is the exciton density. Assuming that $n = 10^{19}$ cm^{-3} (this density was achieved in some experiments,[9]) we obtain $\mu(0) \simeq 2.5$ meV. An ideal gas of excitons with $n = 10^{19}$ cm^{-3} should transform to the Bose-condensed state at $T_c \sim 50$ K, for which $\mu(0)/2T_c \simeq 0.3$.

In the experiment[9] excitons were produced by powerful nanosecond laser pulses at a wavelength $\lambda \simeq 500$ nm focused into the spot of diameter $d \simeq 30$ μm on the sample surface. The volume of the exciton system interacting with the incident light stimulating the two-photon emission can be estimated as $V = d^2 l$, where $l \simeq 1$ μm is the penetration depth of radiation with wavelength 500 nm.

As $\omega \to \Omega_-$ ($\alpha_L \to 0$), the cross section σ^L increases. We write $\hbar(\Omega_- - \omega) = \mu(0)$. In this case

$$\mathbf{f}(\omega_L) \simeq \mathbf{f}(\omega_L') \simeq \mathbf{F}, \quad \mathbf{F} = \frac{1}{4\pi}\int \mathbf{F}^{>}(p_L,k)do_{p_L},$$

where $\mathbf{F}$ is the matrix element of the radiative phonon-assisted recombination of an isolated exciton.[20] This matrix element can be estimated by the formula

$$\frac{1}{\tau_{\mathrm{exc}}} = \frac{4\Omega_-^3}{3c^3\hbar}\mathbf{F}^2, \tag{43}$$

where τ_{exc} is the lifetime of an isolated exciton due to its spontaneous recombination with the emission of a photon of energy $\hbar\Omega_-$ and an optical phonon with energy $\hbar\omega_0^s$. The lifetime of paraexcitons in Cu_2O is $\tau_{\mathrm{exc}} \sim 100$ μs (Ref. 7).

The relaxation time τ^L in the system of Bose-condensed excitons is a subject of further investigation. Even for zero temperature, it can be considerably shorter than the radiative lifetime of excitons τ_{exc} because a quasiparticle can disappear, for example, due to the emission of one or several acoustic phonons. Assuming that the time τ^L is within the interval 10^{-11}–10^{-5} s (the lower bound is defined by the condition $\Gamma_L = 10^{-1}\epsilon(p_L)$, the upper bound is $10^{-1}\tau_{\mathrm{exc}}$), we obtain an estimate for the cross section of stimulated two-photon emission by Bose-condensed paraexcitons in Cu_2O at $T=0$: $\sigma^L = 10^{-16}$–10^{-10} cm^2.

The radiative lifetime of orthoexcitons in Cu_2O is $\tau_{\mathrm{exc}} \sim 300$ ns. Assuming the relaxation time τ^L in a system of orthoexcitons in the Bose-condensed state to be within 10^{-11}–10^{-9} s (in this case the upper bound is determined by the time of transition between the orthoexciton and paraexciton states) yields $\sigma^L = 10^{-11}$–10^{-9} cm^2 at $T=0$. Thus, stimulated two-photon emission accompanied by coherent two-exciton recombination can be experimentally detected in Cu_2O.

The cross section of RLS with coherent two-exciton recombination accompanied by the production of two optical phonons is determined by the squared product of two matrix elements

$$\mathbf{f}'(\tilde{\omega}_L)=\frac{1}{4\pi}\int \mathbf{f}'^{>}(\tilde{p}_L,k)do_{\tilde{p}_L},$$

$$\mathbf{f}(\tilde{\omega}'_L)=\frac{1}{4\pi}\int \mathbf{f}^{>}(\tilde{p}_L,k')do_{\tilde{p}_L},$$

where $\tilde{p}_L$ is determined by the condition $\epsilon(\tilde{p}_L)=\omega+\Omega_-$ (see Sec. 3). The band gap in Cu_2O is wide ($\Omega_-\sim 10^2\omega_0^s$), so $\epsilon(\tilde{p}_L)\gg\omega_0^s$. Using the approach suggested in Appendix A, one can prove that in this case $\mathbf{f}(\tilde{\omega}'_L)$ and $\mathbf{f}'(\tilde{\omega}_L)$ are negligible in comparison with the matrix elements $\mathbf{f}(\omega_L)$ and $\mathbf{f}(\omega'_L)$ in Eq. (42) at $|\Omega_--\omega|\sim\mu(0)$. Moreover, the cross section of the RLS under consideration is proportional to the lifetime of a quasiparticle with energy $\epsilon(\tilde{p}_L)=\omega+\Omega_-$, which is much shorter than the relaxation time τ^L in the cross section (42) for $|\Omega_--\omega|\sim\mu(0)$. Thus, unlike stimulated two-photon emission, one can hardly detect RLS accompanied by coherent two-exciton recombination in Cu_2O. The situation is similar in the case of RLS with coherent two-exciton production.

5. CONCLUSIONS

In this paper, we have demonstrated that coherent two-exciton recombination, i.e., the simultaneous recombination of two excitons with opposite momenta, which corresponds to the existence of nondiagonal long-range order in the system expressed by nonvanishing anomalous averages of the form $\langle N-2|Q_{-p}Q_p|N\rangle$, is possible in a Bose-condensed exciton system interacting with the electromagnetic field. Similarly, coherent two-exciton production corresponding to anomalous averages like $\langle N-2|Q^+_{-p}Q^+_p|N\rangle$ is also possible. In these processes, the exciton occupation numbers are unchanged, and the final state of the exciton system differs from the initial one only in the average number of excitons with zero momentum. Coherent two-exciton recombination may also cause Raman light scattering by excitons in Bose-condensed state (RLS can also be accompanied by coherent two-exciton production). The recoil momentum corresponding to the change in the momentum of electromagnetic field is transferred to phonons or impurities. Both the stimulated two-photon emission and RLS with coherent two-exciton recombination (production) can occur only in the presence of Bose condensate in a system of interacting excitons, so the observation of these effects can be used as a strong experimental evidence for the existence of Bose condensation in exciton systems.

Using diagram methods, we have developed a technique for calculating the cross sections of stimulated two-photon emission and RLS accompanied by coherent two-exciton recombination (or production) at $T>0$. In this approach, the elements of the scattering matrix corresponding to the processes in question are expressed in a natural manner in terms of Green's functions of Bose-condensed excitons [see Eqs. (9), (10), and also (49)].

If the incident light frequency satisfies $\omega<2\Omega_-$ with $\Omega_-=\Omega-\omega_0^s$ (Ω is the frequency of light due to the recombination of an exciton with zero momentum and ω_0^s is the optical phonon frequency), stimulated two-photon emission and RLS accompanied by coherent phonon-assisted two-exciton recombination give rise to a spectral line at frequency $2\Omega_--\omega$ and the anti-Stokes component at $\omega+2\Omega_-$, respectively. For $\omega>2\Omega_-$ the RLS spectrum contains both the anti-Stokes and Stokes components at frequencies $\omega\pm 2\Omega_-$. The anti-Stokes line corresponds due to coherent two-exciton recombination, whereas the Stokes component is due to coherent phonon-assisted two-exciton production. In this case, stimulated two-photon emission is impossible.

Using the approximation (19), we have derived expressions for the cross sections of the processes under consideration at finite temperatures. In the limit $|\omega-\Omega_-|\ll T_c$ (T_c is the temperature of Bose condensation), the cross section of stimulated two-photon emission is a nonmonotonic function of temperature. It increases in a certain temperature range below T_c and can even be larger than it is at $T=0$. The cause of this nonmonotonic behavior is that the cross section of the stimulated two-photon emission is determined not only by the density of excitons in the condensate, which decreases as the temperature increases and vanishes at $T=T_c$, but also by the occupation numbers of quasiparticles with energies $|\omega-\Omega_-|$ in the exciton system, which increases as the temperature grows.

Our estimates indicate that for $|\omega-\Omega_-|\sim\mu(0)$, where $\mu(0)$ is the exciton chemical potential measured with respect to the exciton band bottom, a spectral line at $2\Omega_--\omega$ due to the stimulated two-photon emission accompanied by coherent optical phonon-assisted two-exciton recombination can be experimentally observed in Cu_2O.

The work was supported by grants from INTAS, Russian Fund for Fundamental Research, and Physics of Solid-State Nanostructures program.

APPENDIX

Effective matrix elements of exciton recombination

The objective of this Appendix is to prove that the two-photon emission and RLS accompanied by coherent two-exciton recombination can be analyzed from first principles, without using the effective Hamiltonian (3). Taking as an example two-photon emission, we will determine the conditions under which the analysis based on the effective Hamiltonian (3) is correct. In addition, we will show that the effective matrix elements of exciton recombination used in this paper do not depend on temperature and are equal to those calculated previously for $T=0$ (Ref. 20).

The Hamiltonian describing the interaction between excitons, phonons, and electromagnetic field can be written as

$$\hat{V}(t)=\hat{W}(t)+\hat{D}(t),$$

$$\hat{W}(t)=\sum_{pq}\left[W_{qp}Q^+_q(t)Q_p(t)b_{q-p}(t)+W^*_{pq}Q^+_q(t)Q_p(t)b^+_{p-q}(t)\right],$$

$$\hat{D}(t)=\sum_q [D_q e^{-i\Omega t}Q_q(t)c_q^+(t) + D_q' e^{-i\Omega t}Q_{-q}(t)c_q(t)+\text{H.c.}], \quad (44)$$

where the Hamiltonian $\hat{W}(t)$ describes the scattering of excitons by phonons, $\hat{D}(t)$ is responsible for the interaction between excitons and electromagnetic field, $D_q = i\sqrt{2\pi\omega_q}\mathbf{e}^*\cdot\mathbf{d}_q$, $D_q' = -i\sqrt{2\pi\omega_q}\mathbf{e}\cdot\mathbf{d}_q$.

It is clear that the two-photon emission accompanied by coherent phonon-assisted two-exciton recombination is fourth-order in the Hamiltonian $\hat{V}(t)$. For the element of the S-matrix of the two-photon emission due to coherent two-exciton recombination and production of two optical phonons with momenta $\mathbf{p}-\mathbf{k}$ and $-\mathbf{p}-\mathbf{k}'$ averaged with the Gibbs distribution, we have

$$(S_p)_{fi}=\frac{(-i)^4}{4!}\int_{-\infty}^{\infty}\dots\int \langle f|T_t[\hat{W}(t_1)\hat{W}(t_2)\hat{D}(t_3)\hat{D}(t_4) + \hat{W}(t_1)\hat{D}(t_2)\hat{W}(t_3)\hat{D}(t_4) + \hat{D}(t_1)\hat{W}(t_2)\hat{W}(t_3)\hat{D}(t_4) + \hat{W}(t_1)\hat{D}(t_2)\hat{D}(t_3)\hat{W}(t_4) + \hat{D}(t_1)\hat{W}(t_2)\hat{D}(t_3)\hat{W}(t_4) + \hat{D}(t_1)\hat{D}(t_2)\hat{W}(t_3)\hat{W}(t_4)]|i\rangle dt_1\dots dt_4. \quad (45)$$

Here $\langle f|\dots|i\rangle=\Sigma_n\exp[(F-E_n(N)+\mu N)/T]\langle m|\dots|n\rangle$, where $|n\rangle=|n,N\rangle_{\text{exc}}|i\rangle_{\text{phon}}|i\rangle_{\text{phot}}$ and $|m\rangle=|n,N-2\rangle_{\text{exc}}|f\rangle_{\text{phon}}|f\rangle_{\text{phot}}$, and the other notation is given in Sec. 2.

By changing the time variables in each summand of Eq. (45) we can transform it to

$$(S_p)_{fi}=\frac{1}{4}\int_{-\infty}^{\infty}\dots\int \langle f|T_t\hat{W}(t_1)\hat{W}(t_2) \times\hat{D}(t_3)\hat{D}(t_4)|i\rangle dt_1\dots dt_4, \quad (46)$$

where

$$\begin{aligned}
&\langle f|T_t\hat{W}(t_1)\hat{W}(t_2)\hat{D}(t_3)\hat{D}(t_4)|i\rangle \\
&= D_k D_{k'}\exp[-i\Omega(t_3+t_4)]\sum_{p_1p_2} W_{p_1,p_1+p+k'}W_{p_2,p_2-p+k} \\
&\times\{[\langle T_tQ^+_{p_1+p+k'}(t_1)Q_{p_1}(t_1)Q^+_{p_2-p+k}(t_2)Q_{p_2}(t_2)Q_k(t_3)Q_{k'}(t_4)\rangle\times\langle f|T_tb^+_{-p-k'}(t_1)b^+_{p-k}(t_2)|i\rangle_{\text{phon}} \\
&+\langle T_tQ^+_{p_2-p+k}(t_1)Q_{p_2}(t_1)Q^+_{p_1+p+k'}(t_2)Q_{p_1}(t_2)Q_k(t_3)Q_{k'}(t_4)\rangle\langle f|T_tb^+_{p-k}(t_1)b^+_{-p-k'}(t_2)|i\rangle_{\text{phon}}] \\
&\times\langle f|T_tc^+_k(t_3)c^+_{k'}(t_4)|i\rangle_{\text{phot}}+[\langle T_tQ^+_{p_1+p+k'}(t_1)Q_{p_1}(t_1)Q^+_{p_2-p+k}(t_2)Q_{p_2}(t_2)Q_{k'}(t_3)Q_k(t_4)\rangle \\
&\times\langle f|T_tb^+_{-p-k'}(t_1)b^+_{p-k}(t_2)|i\rangle_{\text{phon}}+\langle T_tQ^+_{p_2-p+k}(t_1)Q_{p_2}(t_1)Q^+_{p_1+p+k'}(t_2)Q_{p_1}(t_2)Q_{k'}(t_3)Q_k(t_4)\rangle \\
&\times\langle f|T_tb^+_{p-k}(t_1)b^+_{-p-k'}(t_2)|i\rangle_{\text{phon}}]\langle f|T_tc^+_{k'}(t_3)c^+_k(t_4)|i\rangle_{\text{phot}}\}. \quad (47)
\end{aligned}$$

Here $\langle\dots\rangle=\Sigma_n\exp[(F-E_n(N)+\mu N)/T]\langle n,N-2|\dots|n,N\rangle_{\text{exc}}$.

In the most interesting case $k\neq k'$, we have

$$\begin{aligned}
&\sum_{p_1p_2} W^*_{p_1,p_1+p+k'}W^*_{p_2,p_2-p+k} \\
&\langle T_tQ^+_{p_1+p+k'}(t_1)Q_{p_1}(t_1)Q^+_{p_2-p+k}(t_2)Q_{p_2}(t_2)Q_k(t_3) \\
&\times Q_{k'}(t_4)\rangle=\sum_{p_1} W^*_{p_1p_1}[W^*_{-k'k}\langle T_tQ^+_{p_1}(t_1)Q_{p_1}(t_1)\rangle\langle T_tQ^+_k(t_2)Q_k(t_3)\rangle\langle T_tQ_{-k'}(t_2)Q_{k'}(t_4)\rangle+W^*_{-kk'}\langle T_tQ^+_{p_1}(t_1)Q_{p_1}(t_1)\rangle \\
&\times\langle T_tQ^+_{k'}(t_2)Q_{k'}(t_4)\rangle\langle T_tQ_{-k}(t_2)Q_k(t_3)\rangle]\delta(p+k')+\sum_{p_2} W^*_{p_2p_2}[W^*_{-k'k}\langle T_tQ^+_{-k}(t_1)Q_k(t_3)\rangle\langle T_tQ_{-k'}(t_1)Q_{k'}(t_4)\rangle \\
&\times\langle T_tQ_{p_2}(t_2)Q^+_{p_2}(t_2)\rangle+W^*_{-kk'}\langle T_tQ^+_{k'}(t_1)Q_{k'}(t_4)\rangle\langle T_tQ_{-k}(t_1)Q_k(t_3)\rangle\langle T_tQ^+_{p_2}(t_2)Q_{p_2}(t_2)\rangle]\delta(p-k) \\
&+W^*_{-k,p-q}W^*_{-k',q-p}\langle T_tQ^+_{p-q}(t_1)Q^+_{q-p}(t_2)\rangle\langle T_tQ_{-k}(t_1)Q_k(t_3)\rangle\langle T_tQ_{-k'}(t_2)Q_{k'}(t_4)\rangle \\
&+W^*_{-k'p}W^*_{-k,-p}\langle T_tQ^+_p(t_1)Q^+_{-p}(t_2)\rangle\langle T_tQ_{-k'}(t_1)Q_{k'}(t_4)\rangle\langle T_tQ_{-k}(t_2)Q_k(t_3)\rangle+W^*_{q-p,k}W^*_{p-q,k'}\langle T_tQ^+_k(t_1)Q_k(t_3)\rangle
\end{aligned}$$

$$\times\langle T_t Q_{q-p}(t_1)Q_{p-q}(t_2)\rangle\langle T_t Q^+_{k'}(t_2)Q_{k'}(t_4)\rangle + W^*_{-pk'}W^*_{pk}\langle T_t Q^+_{k'}(t_1)Q_{k'}(t_4)\rangle\langle T_t Q_{-p}(t_1)Q_p(t_2)\rangle\langle T_t Q^+_k(t_2)Q_k(t_3)\rangle$$
$$+ W^*_{-k,p-q}W^*_{p-q,k'}\langle T_t Q^+_{p-q}(t_1)Q_{p-q}(t_2)\rangle\langle T_t Q_{-k}(t_1)Q_k(t_3)\rangle\langle T_t Q^+_{k'}(t_2)Q_{k'}(t_4)\rangle + W^*_{-k'p}W^*_{pk}\langle T_t Q^+_p(t_1)Q_p(t_2)\rangle$$
$$\times\langle T_t Q_{-k'}(t_1)Q_{k'}(t_4)\rangle\langle T_t Q^+_k(t_2)Q_k(t_3)\rangle + W^*_{q-p,k}W^*_{-k',q-p}\langle T_t Q^+_k(t_1)Q_k(t_3)\rangle\langle T_t Q_{q-p}(t_1)Q^+_{q-p}(t_2)\rangle$$
$$\times\langle T_t Q_{-k'}(t_2)Q_{k'}(t_4)\rangle + W^*_{-pk'}W^*_{-k,-p}\langle T_t Q^+_{k'}(t_1)Q_{k'}(t_4)\rangle\langle T_t Q_{-p}(t_1)Q^+_{-p}(t_2)\rangle\langle T_t Q_{-k}(t_2)Q_k(t_3)\rangle. \quad (48)$$

Similar expressions can be derived from the rest of the summands in Eq. (47).

By substituting Eq. (48) in (46) and performing integration over time variables, we obtain for $\mathbf{p}+\mathbf{k}'\neq 0$ and $\mathbf{p}-\mathbf{k}\neq 0$ (the phonons are assumed to be optical)

$$(S_p)_{fi} = 2\pi i T_{k'k}(\mathbf{p})[(\sqrt{2}-1)\delta(\mathbf{p}-\mathbf{q}/2)+1]\times\delta(\omega'+\omega-2\Omega_-),$$

$$\begin{aligned}T_{k'k}(\mathbf{p}) = D_k D_{k'}[&W^*_{q-p,k}W^*_{p-q,k'}G_k(\omega-\Omega)G_{k'}\\ &\times(\Omega_- -\omega-\omega_0^s)\tilde G_{p-q}(\omega-\Omega_-)\\ &+W^*_{-k,p-q}W^*_{-k',q-p}\tilde G_{-k}(\omega-\Omega)\tilde G_{k'}\\ &\times(\omega-\Omega_-+\omega_0^s)\tilde G^+_{q-p}(\omega-\Omega_-)\\ &+W^*_{q-p,k}W^*_{-k',q-p}G_k(\omega-\Omega)\tilde G_{k'}\\ &\times(\omega-\Omega_-+\omega_0^s)G_{q-p}(\omega-\Omega_-)\\ &+W^*_{-k,p-q}W^*_{p-q,k'}\tilde G_{-k}(\omega-\Omega)G_{k'}(\Omega_- -\omega\\ &-\omega_0^s)G_{p-q}(\Omega_- -\omega)+W^*_{-pk'}W^*_{pk}G_{k'}(\Omega_-\\ &-\omega-\omega_0^s)G_k(\omega-\Omega)\tilde G_{-p}(\omega-\Omega_-)\\ &+W^*_{-k'p}W^*_{-k,-p}\tilde G_{-k}(\omega-\Omega)\tilde G_{k'}(\omega-\Omega_-\\ &+\omega_0^s)\tilde G^+_p(\omega-\Omega_-)+W^*_{-k'p}W^*_{pk}G_k(\omega\\ &-\Omega)\tilde G_{k'}(\omega-\Omega_-+\omega_0^s)G_p(\omega-\Omega_-)\\ &+W^*_{-pk'}W^*_{-k,-p}\tilde G_{-k}(\omega-\Omega)G_{k'}(\Omega_- -\omega\\ &-\omega_0^s)G_{-p}(\Omega_- -\omega)],\end{aligned} \quad (49)$$

where $\tilde G_p(\omega)$, $\tilde G^+_p(\omega)$, and $G_p(\omega)$ are Fourier transforms of the anomalous and normal Green's functions of excitons in the Bose-condensed state, which are defined as follows:

$$G_p(t-t') = -i\langle T_t Q_p(t)Q^+_p(t')\rangle,$$
$$\tilde G^{(+)}_p(t-t') = -i\langle T_t Q^{(+)}_{-p}(t)Q^{(+)}_p(t')\rangle. \quad (50)$$

Thus, we have derived the expression for the element of the *S*-matrix responsible for the two-photon emission due to coherent recombination directly from the Hamiltonian of the interaction between excitons and electromagnetic field and the Hamiltonian of the exciton–phonon interaction (44). Elements of the *S*-matrix corresponding to the RLS accompanied by coherent two-exciton recombination (or production) can be derived similarly.

In the general case, Eq. (49) cannot be reduced to the corresponding expression (12) derived from the effective Hamiltonian (3). Below we will determine the conditions under which this is possible and derive an expression for $L^>_{pq}$.

By analyzing stimulated two-photon emission under the condition $|\omega-\Omega_-|\ll\omega_0^s$, we obtain

$$\begin{aligned}T_{k'k}(\mathbf{p}) = D_k D_{k'}[&W^*_{q-p,k}W^*_{p-q,k'}G_k(-\omega_0^s)G_{k'}\\ &\times(-\omega_0^s)\tilde G_{p-q}(\omega-\Omega_-)\\ &+W^*_{-k,p-q}W^*_{-k',q-p}\tilde G_{-k}\\ &\times(-\omega_0^s)\tilde G_{k'}(\omega_0^s)\tilde G^+_{q-p}(\omega-\Omega_-)\\ &+W^*_{q-p,k}W^*_{-k',q-p}G_k(-\omega_0^s)\tilde G_{k'}(\omega_0^s)G_{q-p}(\omega\\ &-\Omega_-)+W^*_{-k,p-q}W^*_{p-q,k'}\tilde G_{-k}(-\omega_0^s)G_{k'}\\ &\times(-\omega_0^s)G_{p-q}(\Omega_- -\omega)+W^*_{-pk'}W^*_{pk}G_{k'}\\ &\times(-\omega_0^s)G_k(-\omega_0^s)\tilde G_{-p}(\omega-\Omega_-)\\ &+W^*_{-k'p}W^*_{-k,-p}\tilde G_{-k}(-\omega_0^s)\tilde G_{k'}(\omega_0^s)\tilde G^+_p(\omega\\ &-\Omega_-)+W^*_{-k'p}W^*_{pk}G_k(-\omega_0^s)\tilde G_{k'}(\omega_0^s)G_p(\omega\\ &-\Omega_-)+W^*_{-pk'}W^*_{-k,-p}\tilde G_{-k}(-\omega_0^s)G_{k'}\\ &\times(-\omega_0^s)G_{-p}(\Omega_- -\omega)].\end{aligned} \quad (51)$$

Under the conditions of approximation (19), we have for the retarded Green's functions

$$G^R_p(\omega) = -2\pi i n_0(T)\delta_p\delta(\omega)+G'^R_p(\omega),$$
$$G'^R_p(\omega) = (1-\delta_p)\frac{\omega+\xi_p}{(\omega-\epsilon_p+i\Gamma_p/2)(\omega+\epsilon_p+i\Gamma_p/2)}, \quad (52)$$

from which an expression for $G_p(\omega)$ can be obtained. The anomalous Green's function $\tilde G_p(\omega)$, which is defined by Eq. (50), is related to the Green's function $\hat G_p(\omega)$ [see definition (8), and Eqs. (15) and (21)] by the formula $\hat G_p(\omega) = -2\pi i n_0(T)\delta_p\delta(\omega)+\hat G_p(\omega)$.

By comparingthe normal and anomalous Green's functions at $\omega=\omega_0^s$, we obtain $\tilde G_k(\omega_0^s)/G_k(\omega_0^s)\ll 1$ for $\omega_0^s \gg \xi_k$. In this case the element (51) of the scattering matrix is determined by the expression

$$T_{k'k}(\mathbf{p})=D_kD_{k'}[W^*_{q-p,k}W^*_{p-q,k'}G_k(-\omega_0^s)G_{k'}\times(-\omega_0^s)\tilde{G}_{p-q}(\omega-\Omega_-)+W^*_{-pk'}W^*_{pk}G_{k'}\times(-\omega_0^s)G_k(-\omega_0^s)\tilde{G}_{-p}(\omega-\Omega_-)], \quad (53)$$

where $G_{k'}(\omega_0^s)\simeq G_k(\omega_0^s)\simeq 1/\omega_0^s$.

The comparison between the latter expression and Eq. (12) in Sec. 2 yields $L^>_{pq}$ for the effective Hamiltonian of exciton recombination:

$$L^>_{pq}=-i\frac{D_qW^*_{pq}}{\omega_0^s}. \quad (54)$$

The expressions for the other matrix elements of the exciton recombination in the effective Hamiltonian (3) can be derived similarly.

Thus, the two-photon emission accompanied by coherent two-exciton recombination can be analyzed with the aid of the effective Hamiltonian (3) with $L^>_{pq}$ determined by Eq. (54) if the incident light satisfies the conditions $|\omega-\Omega_-|\ll\omega_0^s$ and $\xi_k\ll\omega_0^s$. In this case, the matrix element (54) does not depend on temperature and $\mathbf{f}(\omega_L)$ is identical to the effective matrix element $\mathbf{F}$ responsible for the phonon-assisted recombination of an isolated exciton.[20]

APPENDIX B

Luminescence intensity of Bose-condensed excitons

The luminescence of Bose-condensed excitons for frequency $\omega'<\Omega_-$ is due to the optical phonon-assisted exciton recombination accompanied by the production of a Bogoliubov quasiparticle with energy $\epsilon(p'_L)=\Omega_--\omega'$ in the exciton system. The matrix element of this recombination is

$$L=L^>_{p'_Lk}v_{p_L}\sqrt{n_{p'_L}+1},\quad v^2_{p'_L}=\frac{\sqrt{\alpha'^2_L+1}-\alpha'_L}{2\alpha'_L},$$
$$n_{p'_L}=\left[\exp\left(\frac{|\Delta\omega'|}{T}\right)-1\right]^{-1}, \quad (55)$$

where $\Delta\omega'=\omega'-\Omega_-$ and $\alpha'_L=|\Delta\omega'|/\zeta(T)$.

Using the Fermi ''golden rule,'' we obtain the optical phonon-assisted luminescence intensity $I^L_s(\omega')$ of Bose-condensed excitons at frequencies $\omega'<\Omega_-$:

$$I^L_s(\omega')=\frac{\omega'^4\sqrt{2m^3\zeta(T)(\sqrt{\alpha'^2_L+1}-1)}\mathbf{f}^2(\omega'_L)}{6\pi^2c^3\sqrt{\alpha'^2_L+1}(\sqrt{\alpha'^2_L+1}+\alpha'_L)}\times\left[1+\coth\frac{|\Delta\omega'|}{2T}\right]. \quad (56)$$

The luminescence of Bose-condensed excitons for a frequency $\omega'>\Omega_-$ is due to the optical phonon-assisted exciton recombination accompanied by the disappearance of a quasiparticle with energy $\epsilon(p'_L)=\omega'-\Omega_-$. The matrix element of this recombination is derived from Eq. (55) by substituting $v_{p'_L}\sqrt{n_{p'_L}+1}\to\sqrt{1+v^2_{p'_L}}\sqrt{n_{p'_L}}$. The luminescence intensity at frequency $\omega'>\Omega_-$ is derived from Eq. (56) by substituting $\alpha'_L\to-\alpha'_L$ and $\coth(|\Delta\omega'|/2T)+1\to\coth(|\Delta\omega'|/2T)-1$. Thus, the expression that determines the luminescence spectrum of Bose-condensed excitons at an arbitrary frequency $\omega'\neq\Omega_-$ has the form

$$I^L_s(\omega')=\frac{\omega'^4\sqrt{2m^3\zeta(T)(\sqrt{\alpha'^2_l+1}-1)}\mathbf{f}^2(\omega'_L)}{6\pi^2c^3\sqrt{\alpha'^2_L+1}[\sqrt{\alpha'^2_L+1}-\mathrm{sign}(\Delta\omega')\alpha'_L]}\times\left[\mathrm{sign}(-\Delta\omega')+\coth\frac{|\Delta\omega'|}{2T}\right]. \quad (57)$$

[a)]E-mail: lozovik@isan.troitsk.ru

[1)]In the general case, the radiative recombination of an exciton can result in emission (absorption) of an arbitrary number of phonons. When using the Hamiltonian (3), we limit our analysis for simplicity to the case of exciton recombination with emission (absorption) of one phonon.

[2)]Calculations involving the two-photon emission and RLS under discussion could be performed using Keldysh's elegant diagram technique (see, e.g., Ref. 21, which is devoted to a problem that requires a similar technique). In our opinion, however, our approach used in this specific case is more transparent.

[3)]In the general case, the number of phonons involved in the process can be arbitrary. Moreover, the recoil momentum (the whole or a fraction of it) can be transferred to impurities. Thus, stimulated two-photon emission can result in the appearance of the spectral lines at frequencies $2(\Omega-n\omega_0^s)-\omega$, where n is an arbitrary integer.

[4)]In the general case, RLS, like the two-photon emission, can involve an arbitrary number of phonons. Moreover, the recoil momentum can be transferred (entirely or partially) to impurities. Thus, RLS accompanied by coherent two-exciton production or recombination can give rise to anti-Stokes and Stokes components at frequencies $\omega+(2\Omega-n\omega_0^s)=\omega'$ and $\omega-(2\Omega-n\omega_0^s)=\omega'$, respectively, where n is an arbitrary integer (see also Ref. 20).

[1] J. M. Blatt, K. W. Böer, and W. Brandt, Phys. Rev. **126**, 1691 (1962).
[2] S. A. Moskalenko, Fiz. Tverd. Tela **4**, 2770 (1962) [Sov. Phys. Solid State **4**, 2032 (1963)].
[3] V. A. Gergel', R. F. Kazarinov, and R. A. Suris, Zh. Éksp. Teor. Fiz. **53**, 544 (1967) [Sov. Phys. JETP **26**, 354 (1968)].
[4] L. V. Keldysh and A. N. Kozlov, Zh. Éksp. Teor. Fiz. **54**, 978 (1968) [Sov. Phys. JETP **27**, 521 (1968)].
[5] E. Hanamura and H. Haug, Phys. Rep. **33**, 209 (1977).
[6] C. Comte and P. Nozières, J. de Phys. **43**, 1069 (1982).
[7] *Bose–Einstein Condensation*, ed. by A. Griffin, D. W. Snoke, and S. Stringari, Cambridge University Press (1995).
[8] D. W. Snoke, J. P. Wolfe, and A. Mysyrowicz, Phys. Rev. Lett. **64**, 2543 (1990).
[9] Jia Ling Lin and J. P. Wolfe, Phys. Rev. Lett. **71**, 1222 (1993).
[10] E. Fortin, S. Fafard, and A. Mysyrowicz, Phys. Rev. Lett. **70**, 3951 (1993).
[11] E. Benson, E. Fortin, and A. Mysyrowicz, Solid State Commun. **101**, 313 (1997).
[12] C. Ell, A. L. Ivanov, and H. Haug, Phys. Rev. B **57**, 9663 (1998).
[13] A. E. Bulatov and S. G. Tikhodeev, Phys. Rev. B **46**, 15058 (1992).
[14] G. A. Kopelevich, S. G. Tikhodeev, and N. A. Gippius, Zh. Éksp. Teor. Fiz. **109**, 2189 (1996) [JETP **82**, 1180 (1996)].
[15] L. V. Butov, A. Zrenner, G. Abstreiter, G. Bohm, and G. Weimann, Phys. Rev. Lett. **73**, 304 (1994).
[16] Yu. E. Lozovik and V. I. Yudson, JETP Lett. **22**, 274 (1975); Zh. Éksp. Teor. Fiz. **71**, 738 (1976) [Sov. Phys. JETP **44**, 389 (1976)].
[17] Yu. E. Lozovik and O. L. Berman, Zh. Éksp. Teor. Fiz. **97**, 1879 (1997) [JETP **84**, 1027 (1997)].
[18] Yu. E. Lozovik, O. L. Berman, and V. G. Tsvetus, JETP Lett. **66**, 355 (1997); Yu. E. Lozovik, O. L. Berman, and V. G. Tsvetus, submitted to Phys. Rev. B.
[19] Yu. E. Lozovik and A. V. Poushnov, Solid State Commun. **105**, 527 (1998).
[20] Yu. E. Lozovik and A. V. Poushnov, Phys. Rev. B **58**, 6608 (1998).
[21] L. V. Keldysh and S. G. Tikhodeev, Zh. Éksp. Teor. Fiz. **90**, 1852 (1986) [Sov. Phys. JETP **63**, 2086 (1986)].

[22] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, Pergamon Press, Oxford (1982).
[23] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Mechanics*, Dover Publications, New York (1975) [Russian original, Fizmatgiz, Moscow (1962)].
[24] V. N. Popov, *Functional Integrals in Quantum Field Theory and Statistical Physics*, D. Reidel, Dordrecht; Kluwer Academic, Boston (1983) [Russian original, Gosatomizdat, Moscow (1982)].
[25] A. A. Abrikosov and L. A. Fal'kovskiĭ, Zh. Éksp. Teor. Fiz. **40**, 262 (1961) [Sov. Phys. JETP **13**, 179 (1961)].
[26] L. V. Keldysh and Yu. V. Kopaev, Fiz. Tverd. Tela **6**, 2791 (1964) [Sov. Phys. Solid State **6**, 2219 (1965)].
[27] Yu. V. Kopaev, Trudy FIAN **86**, 3 (1975).
[28] H. Shi, G. Verechaka, and A. Griffin, Phys. Rev. B **50**, 1119 (1994).

Translation provided by the Russian Editorial office.

Dynamics of domain walls with drifting Bloch lines in single crystals of yttrium iron garnet

L. M. Dedukh*) and Yu. P. Kabanov

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

(Submitted 23 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1377–1385 (April 1999)

Induction and magnetic methods are used to study the effect of drifting Bloch lines on the wall velocity in a single crystal sample of yttrium iron garnet cut in the form of a long prism with only one 180-degree domain wall. A sharp increase in the velocity and in resonance bending vibrations of the wall are observed when Bloch line drift is initiated. The character of the wall motion is investigated under these conditions. An analysis of the experimental data shows that the effective reduction in the influence of drifting Bloch lines on the characteristics of the wall motion may be related to a magnetic aftereffect phenomenon. © *1999 American Institute of Physics.* [S1063-7761(99)01504-8]

1. INTRODUCTION

It is well known that in the overwhelming majority of ferromagnetic crystals, the Bloch (Néel) lines that develop under static[1] and dynamic[2] conditions under the influence of the magnetostatic stray fields that exist at the sample surfaces are an obligatory structural element of the domain walls. They separate regions with opposite directions of the spin in the domain wall, and are therefore characterized by a vortical spin distribution[3] and distinctive properties that change the character of the motion of a monopolar domain wall significantly. In particular, Bloch lines greatly reduce the mobility[4] and amplitude of the bending eigenmodes[5] of a monopolar wall, and also increase its effective mass.[6]

Direct experimental studies of the behavior of Bloch lines in moving domain walls in single crystal yttrium iron garnet[7] have made it possible to study the motion of Bloch lines affected by gyrotropic forces, to determine their dynamic characteristics (mobility and the longitudinal and transverse components of the effective mass), and their behavior as a function of the amplitude of the exciting field. It has been shown that only in a relatively weak variable magnetic field that excites wall vibrations do the Bloch lines oscillate near their equilibrium positions. When the field exceeds a critical value (H_{cr}), the entire system of Bloch lines in a wall drifts along it at a velocity of several meters per second.[8] The character of the motion of the wall itself changes when Bloch line drift sets in. In particular, the free damped oscillations of the wall initiated by magnetic field pulses and by the oscillations of the lines themselves are replaced by aperiodic damping.[7]

Line drift in yttrium iron garnet is accompanied by intense creation and annihilation of lines, and therefore it would seem that it should affect the velocity of a domain wall in the same way as observed during dynamic transformations of the domain wall structure in highly anisotropic garnet films[2] in which Bloch lines are created, move, and vanish. The first, recent studies of bending vibrations of walls with drifting Bloch lines in yttrium iron garnet show, however, that when Bloch line drift is excited, the effect of the lines on the velocity of a vibrating wall is sharply reduced.[9] Here we describe the results of a comprehensive study of this unusual behavior in domain walls with drifting Bloch lines.

2. EXPERIMENTAL TECHNIQUE

These studies were conducted on a single crystal sample of yttrium iron garnet cut in the form of a $3.2\times0.7\times0.03$ mm^3 rectangular prism. It contained a single 180° domain wall that separated domains magnetized along the $\langle 111\rangle$ directions in the (112) plane of the wafer. In the initial state the wall contained vertical Bloch lines. As necessary, a monopolar state of the wall was created using a sinusoidal field (H_x) parallel to the magnetization in the domains, and which initiated drift of the Bloch lines in the presence of an additional constant field (H_z) perpendicular to the plane of the wafer.[8] The magnetic field was produced by Helmholtz coils with a radius of 6 mm.

The motion of the domain wall was recorded by a small compensated coil wound directly on the sample. The induction signal from this coil was fed to an SK4-59 spectrum analyzer (to record the spectra of the vibrations), an S8-9 storage oscilloscope (to record single pulses), or a V3-39 millivoltmeter (to record the amplitude of the oscillations in the wall velocity). Motion of the Bloch lines along the wall was studied by a magnetooptical method using a polarizing microscope equipped with a photomultiplier.[8] The signals were recorded and processed, and control of the measurement instruments was coordinated and synchronized by a personal computer.

3. EXPERIMENTAL RESULTS

Figure 1 shows flexural vibration spectra of the 180° domain wall for various amplitudes (H_{x0}) of the driving sinusoidal field H_x. In a weak exciting field, the spectrum of

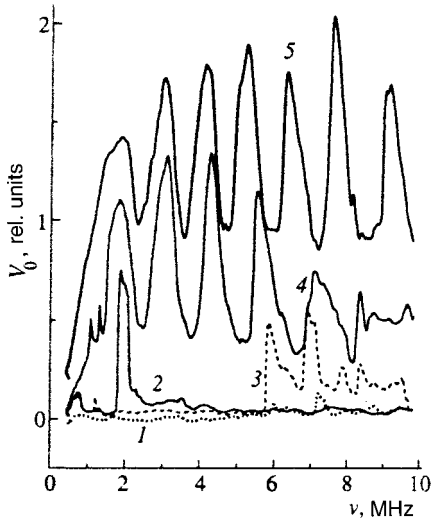

FIG. 1. The induction signal (V_0), proportional to the amplitude of the oscillations in the displacement velocity of a demagnetized 180° domain wall, as a function of the frequency ν of the sinusoidal field H_x for various field amplitudes: $H_{x0}=12$ mOe (*1*), 22.5 (*2,3*), 30 (*4*), and 45 (*5*).

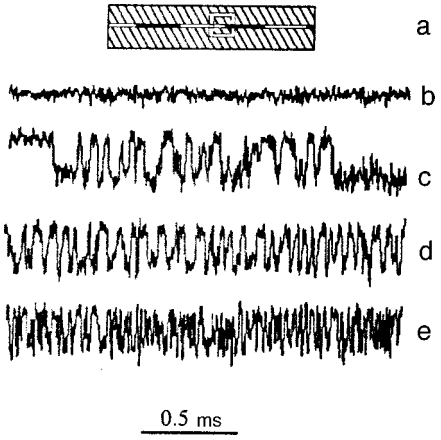

FIG. 2. Magnetooptic imaging of a 180° domain wall in a polarization microscope with an isolated cross section for a photometric measurement (a) and single oscilloscope traces (b–e) characterizing the variation in the intensity of the light transmitted through the photometric cross section with exciting field amplitudes $H_{x0}=15$ mOe (b), 28 (c), 45 (d), and 75 (e). The frequency of the field is $\nu=1.8$ MHz. The change in the intensity of the light is determined by the movement along the wall of ''dark'' and ''light'' subdomains and of nonlinear perturbation s in the spin system in these subdomains.

the flexural oscillations of the wall is essentially invisible in the frequency dependence of the induction signal, which is proportional to the amplitude of the velocity oscillations of the wall ($V_0=\omega y_0=2\pi\nu y_0$, where ν is the frequency of the field and y_0 is the amplitude of the wall vibrations) (curve *1*). As H_{x0} approaches the critical value H_{cr}, isolated peaks show up in the $V_0(\nu)$ curves (curve *2*), which reflect the excitation of flexural vibrations in the wall with wave vectors perpendicular to the domain magnetizations. These peaks could disappear and reappear at the same or different resonance frequencies in successive measurements (curve *3*). When the field reaches H_{cr}, distinct and stable peaks show up in the $V_0(\nu)$ curves at low frequencies (curve *4*), suggesting the excitation of flexural oscillations in the demagnetized wall. High-frequency resonance peaks have not yet formed in curve *4*; these are stabilized as H_{x0} is increased further (curve *5*). In addition, a comparison of curves *4* and *5* shows that as H_{x0} is increased, the resulting resonances shift to lower frequencies.

Figure 2a shows a magnetooptical image of a 180° domain wall in the polarization microscope with slightly uncrossed polarizers. The white and black sections of the wall are characterized by opposite directions of the spin turn, and are separated by vertical Bloch lines. Figures 2b–2d show single oscilloscope traces of the time variation in the intensity of the light, measured with a photomultiplier, in the local part of the crystal indicated by a square in Fig. 2a. These were recorded in the intervals between successive increases in H_{x0} at a frequency of 1.8 MHz, corresponding to one of the resonance peaks in Fig. 1. At low field amplitudes the oscilloscope traces only contained a noise signal (Fig. 2b). The oscillations of the lines along the oscillating wall produced by gyrotropic forces[7] do not show up in the single scope traces. As the field amplitude approached critical, at first single narrow peaks associated with the excitation of solitary nonlinear waves in the domain wall appeared that were similar to those observed in monopolar walls.[8] The oscilloscope traces of Fig. 2c records the passage of several subdomains through the photometrically observed segment. The number of subdomains was different in successive measurements. Unstable peaks showed up in the $V_0(\nu)$ curve at this field level, which initiates unstable drift of the Bloch lines. They emerged only at the onset of line drift, and at those frequencies where that drift occurred. Stable drift of Bloch lines was observed only in the trace of Fig. 2d, which was recorded at a field amplitude for which flexural oscillations of the wall showed up clearly in the $V_0(\nu)$ curve (Fig. 1, curve *5*). Further increases in H_{x0} were accompanied by a reduction in the average period of the magnetooptical signal (Fig. 2e) owing to the increased drift velocity of the Bloch lines.

Thus, a comparison of the data in Figs. 1 and 2 shows that when Bloch line drift is excited, there is a large increase in the displacement velocity and amplitude of flexural oscillations of the domain wall.

Figure 3 shows the amplitude V_0 of the oscillations in the velocity of the wall as a function of the amplitude H_{x0} of the exciting field measured at a resonant frequency of 1.8 MHz. The behavior of the Bloch lines in the wall observed during the time this curve was recorded corresponds to the traces of Fig. 2. The relationship between the wall displacement velocity and the behavior of the Bloch lines in the wall shows up more clearly in this curve. Just when line drift is excited, there is a sharp rise in the displacement velocity of the wall. Furthermore, this curve manifests ''hysteresis'' in the discontinuity in the wall velocity as the field amplitude is raised or lowered. Figure 3 also shows fragments of single magnetooptical oscilloscope traces recorded before and after the velocity jump during forward and reverse scans. These show that the velocity jumps took place during forward or reverse scans of $V_0(H_{x0})$, when the drift of the Bloch lines either set in or ended, respectively, and as

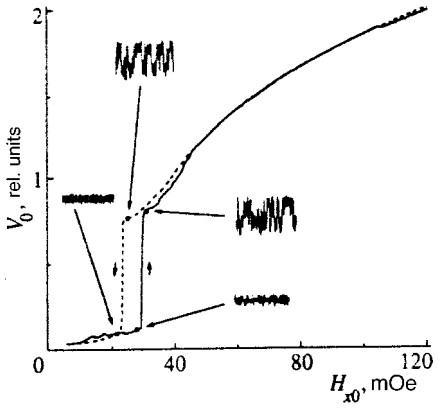

FIG. 3. Amplitude V_0 of the velocity of wall oscillations as a function of the amplitude H_{x0} of an exciting field at $\nu = 1.8$ MHz, and fragments of single-sweep oscilloscope traces of the magnetooptical signal similar to those shown in Fig. 2.

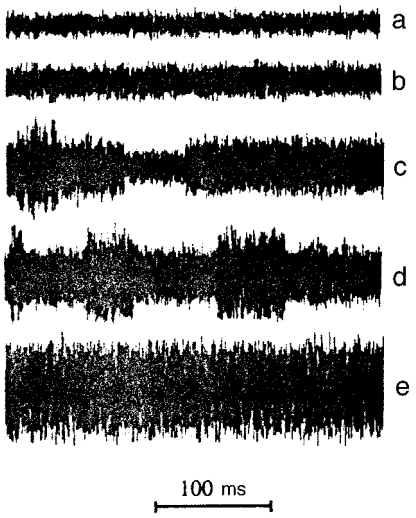

FIG. 5. Single-sweep oscilloscope traces of the induction signal, proportional to the amplitude of the oscillations in the displacement velocity of the wall, which represent the wall oscillations in real time at $\nu = 1$ MHz for field amplitudes $H_{x0} = 15$ mOe (a), 27 (b,c), 28.5 (d), 30 (e).

can be seen in Fig. 3, when the field amplitude is reduced, line drift is observed in the wall at substantially lower fields than when the field H_{x0} is increased.

A comparison of $V_0(H_{x0})$ curves measured with rising field amplitude at the resonant frequencies (Fig. 4, curves *1* and *2*) shows that the critical field H_{cr} at which the velocity jump occurs lies in the range 27–32 mOe and changes little in successive measurements. When these curves were plotted with decreasing H_{x0}, this range extended to lower fields (22 mOe) as a result of the hysteresis . At arbitrary frequencies (Fig. 4, curves *3–5*), the displacement velocity of a wall does not increase so rapidly. This is because during forced oscillations of a wall, stable drift of Bloch lines is initiated more gradually as H_{x0} increases: it first occurs in isolated segments of the wall, and then over the entire wall. In the experiment, a tendency was observed for this transition in the field to widen as the frequency decreased. The fact that when the wall oscillates resonantly, stable line drift occurs over a narrower range of H_{x0} probably suggests that line drift depends directly on the dynamical state of the wall, but not on the magnitude of H_{x0}.

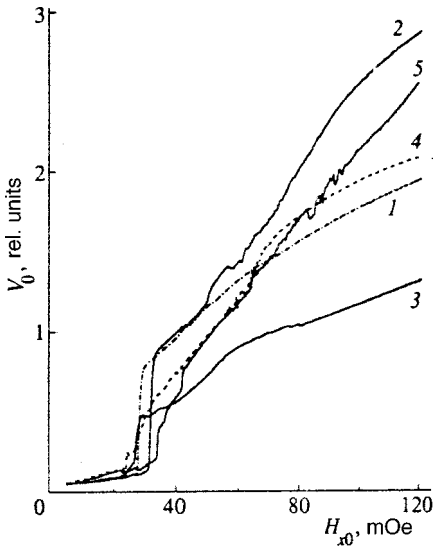

FIG. 4. $V_0(H_{x0})$ curves for resonant (*1*,*2*) and arbitrary (*3–5*) frequencies of the exciting field H_x: $\nu =$ 1.8 MHz (*1*), 3 (*2*), 1 (*3*), 2.3 (*4*), and 3.8 (*5*).

The single-sweep oscilloscope traces of Fig. 5 reflect the amplitude (or velocity V_0) of wall vibrations in near-critical fields at a frequency of 1 MHz. They show that in a weak field, the amplitude y_0 of wall vibrations does not vary in time (Fig. 5a). As the field amplitude approaches the critical value, the oscilloscope traces contain a time interval in which the wall vibrates with a higher amplitude (cf. Figs. 5b and 5c); they increase in Fig. 5c. Intervals with wall vibrations of even greater amplitude are visible in Fig. 5d, which increase as H_{x0} rises further, and finally the wall undergoes a transition to high-velocity motion (Fig. 5e). At the resonant frequencies this transition from ''slow'' to ''fast'' wall movement takes place over a narrower range of H_{x0}. Thus, these data show that at field amplitudes close to H_{cr}, wall motion is manifestly nonstationary, owing to unstable Bloch line drift processes. Wall motion stabilizes somewhat at higher H_{x0}, but it is still not harmonic. The latter may to a large extent be due to the nonuniform character of the Bloch line drift.

Curves *1* and *2* in Fig. 6 represent the measured $V_0(H_{x0})$ curves for a wall in an initially monopolar state at a nonresonant frequency of 0.8 MHz for the exciting field H_x. They were recorded in the presence of additional constant fields of $H_z = 46$ and 40 Oe, respectively, perpendicular to the plane of the wafer, which maintained the monopolar state of the wall. These curves are qualitatively consistent with similar curves from Ref. 10, but they do differ slightly, since they were measured at different ν and H_z. A comparison of curves *1* and *2*, in particular, reveals how the auxiliary field H_z affects the $V_0(H_{x0})$ curve. Clearly, as H_z decreases, the knee in $V_0(H_{x0})$, which characterizes the onset of nonlinear excitations in the wall,[8] moves to lower amplitudes of the exciting field.

Curve *3* in Fig. 6 was recorded for walls containing

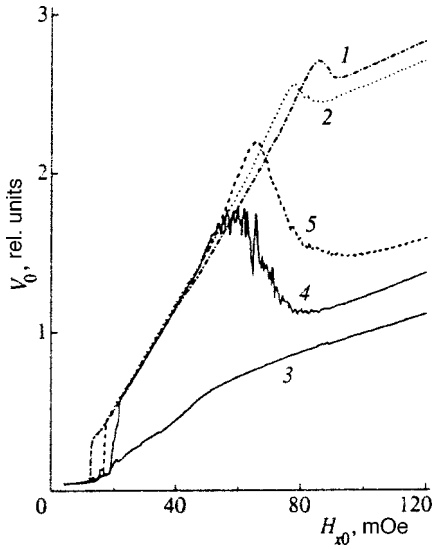

FIG. 6. $V_0(H_{x0})$ at 0.8 MHz for domain walls in monopolar (*1,2*) and demagnetized (*3*–5) initial states in the presence of an additional field $H_z = 46$ Oe (*1*), 40 (*2*), 20 (*5*), 16 (*4*), and 0 (*3*).

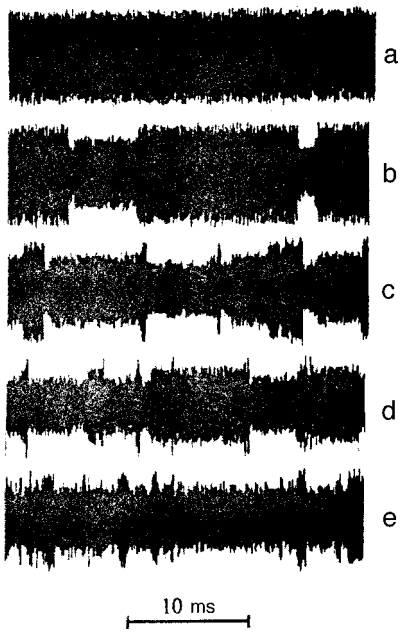

FIG. 7. Single-sweep oscilloscope traces of the induction signal representing wall oscillations in real time under the conditions of curve *4* of Fig. 6 in the region where V_0 is falling: $H_{x0} = 39$ mOe (a), 42 (b), 43.5 (c), 48 (d), and 67.5 (e); $H_z = 16$ Oe, $\nu = 0.8$ MHz.

Bloch lines. $V_0(H_{x0})$ for a demagnetized wall clearly falls below curves *1* and *2*, and it has essentially no sharp rise in V_0 owing to the excitation of line drift, since it was recorded at a low, nonresonant frequency of H_x. If the crystal is placed in an additional low field $H_z = 16$ Oe , then $V_0(H_{x0})$ for the demagnetized wall takes the form of curve *4* in Fig. 6. Direct observations of the behavior of the Bloch lines and measurements of single magnetooptical oscilloscope traces similar to those presented above showed that the sudden rise in the wall velocity in curve *4* near $H_{x0} = 20$ mOe was caused by excitation of Bloch lines in the wall, which in the presence of the field H_z led to formation of a monopolar state in the wall. Curve *4* therefore coincided with curves *1* and *2*. However, when the amplitude of the exciting field is increased further, a nonmonotonic reduction in V_0 is observed in curve *4* owing to the excitation of nonlinear perturbations in the wall that disrupt the monopolar state, and ultimately lead to the appearance of drifting lines in the wall. The wall velocity therefore decreases to a minimum, after which it rises slightly.

It is typical that in this range of H_{x0}, curve *4*, which characterizes the velocity of a wall with drifting lines, is not coincident with curve *3*. This is because the flexural vibrations of a demagnetized wall (curve *3*) are weaker than the vibrations of a ''magnetized'' wall (curve *4*).[11] As the field H_z increased, a monopolar state developed in the wall for a weaker field H_x and was maintained over a wider range of its amplitude, so that the drop in V_0 took place at a higher exciting field (curve *5*). We note that similar behavior of vibrating walls has also been observed when an additional constant field (H_y) is applied perpendicular to the plane of the wall. It enhanced the drift of Bloch lines[12], and therefore stimulated the formation of a monopolar state in the wall over a certain range of H_{x0} that depends on H_y.

Figure 7 shows single oscilloscope traces representing oscillations of the wall over time under the conditions for curve *4* of Fig. 6 in the neighborhood of the field H_x where V_0 drops. These show a gradual transition from ''fast'' to ''slow'' wall motion, in complete agreement with curve *4* (Fig. 6). In addition, these traces reveal the nonstationary nature of the wall motion associated with the unstable excitation of nonlinear processes in a monopolar wall, which leads to the formation of drifting Bloch lines in this wall.

4. DISCUSSION

The data presented here show that the excitation of Bloch line drift leads to a rise in the velocity of a domain wall containing Bloch lines, with the greatest enhancement in the wall velocity being observed at resonant frequencies. This sort of behavior of the domain wall is most likely related to a magnetic aftereffect.[13] In our case, this phenomenon is determined by the interaction of Bloch lines and domain walls with point defects of the crystal lattice, whose state depends on the direction of the spins in them. According to the theory of this phenomenon, when either domain walls[13] or Bloch lines[14] interact with point defects they create a potential surface for themselves, and the greater the amplitude of their vibrations and their velocity, the lower the height of the relief. When they drift, the Bloch lines, first of all, cannot create a potential surface for their motion, so their influence on the motion of the domain wall should be weakened. In addition, every displacement of a Bloch line along a domain wall reverses the sense of spin precession in the wall, and thereby reduces the relief produced by the spins in the wall itself.

This has been partially confirmed in Ref. 9, where it is shown that wall flexural vibrations are enhanced when an additional magnetic field that induces low-frequency oscillations of Bloch lines along the wall is applied to a crystal.

Another important argument in favor of this explanation is the observed hysteresis in the wall velocity jump (see Fig. 3), which is analogous to the hysteresis in $V_0(H_{x0})$ measured at a monopolar wall in yttrium iron garnet. The latter has been explained in terms of a magnetic aftereffect.[10] In our case, when the field amplitude increases, Bloch line drift is initiated in a high field (compared to the optimum), since the Bloch lines must overcome the potential relief that they produce at a lower H_{x0}. When H_{x0} is reduced, drifting Bloch lines inhibit the formation of potential relief near the Bloch lines, so that drift of the lines is cut off in a lower field than the field that initiated their drift. Thus, in a backward scan of the $V_0(H_{x0})$ curve, Bloch line drift occurs at much lower fields than in a forward scan (see the oscilloscope traces of Fig. 3).

In conclusion, we also note that Bloch line drift is excited in yttrium iron garnet in very weak exciting fields. $H_{\rm cr}$ is considerably weaker than the Walker field $H_W = 2\pi M\alpha \approx 60$ mOe (where α is a dimensionless parameter characterizing the viscosity and M is the saturation magnetization), which characterizes the onset of a dynamic transformation of the wall structure associated with precession of the spins in the wall, and thereby leads to time-dependent motion of the wall.[15] Thus, the present experiments show that time-dependent motion of the domain walls occurs in fields well below H_W.

Also noteworthy is the experimentally observed feasibility of deliberately altering the velocity of a vibrating wall (see Fig. 6), and thereby the high-frequency permeability of yttrium iron garnet, which may be useful in certain applications.

This work was supported by the Russian Fund for Fundamental Research (Grant no. 97-02-16879).

*)E-mail: dedukh@issp.ac.ru

[1] S. Shtrikman and D. Treves, J. Appl. Phys. **31**, 375, 1304 (1960).
[2] A. Malozemov and J. Slonczewski, *Domain Walls in Bubble Materials*, Academic Press, New York (1979).
[3] A. V. Nikiforov and É. B. Sonin, JETP Lett. **40**, 1119 (1984).
[4] J. C. Slonczewski, Phys. Rev. Lett. **29**, 1679 (1972); A. P. Malozemov and J. C. Slonczewski, Phys. Rev. Lett. **29**, 952 (1972); T. M. Morris, G. J. Zimmer, and F. B. Humphrey, J. Appl. Phys. **47**, 721 (1976).
[5] L. M. Dedukh and Yu. P. Kabanov, J. Magn. Magn. Mater. **147**, 355 (1995).
[6] W. Jantz, J. D. Slonczewski, and B. E. Argyle, J. Magn. Magn. Mater. **23**, 8 (1981).
[7] V. S. Gornakov, L. M. Dedukh, V. I. Nikitenko, and V. T. Synogach, Zh. Éksp. Teor. Fiz. **90**, 2090 (1986) [Sov. Phys. JETP **63**, 1225 (1986)].
[8] V. S. Gornakov, L. M. Dedukh, and V. I. Nikitenko, Zh. Éksp. Teor. Fiz. **86**, 1505 (1984) [Sov. Phys. JETP **59**, 881 (1984)].
[9] A. B. Shumm, L. M. Dedukh, and Yu. P. Kabanov, JETP Lett. **67**, 78 (1998).
[10] V. S. Gornakov, V. I. Nikitenko, I. A. Prudnikov, and V. T. Synogach, Phys. Rev. B **46**, 10829 (1992).
[11] L. M. Dedukh, V. I. Nikitenko, and V. T. Synogach, Acta Phys. Pol. A **76**, 295 (1989).
[12] V. S. Gornakov, L. M. Dedukh, and V. I. Nikitenko, Zh. Éksp. Teor. Fiz. **94**, 245 (1988) [Sov. Phys. JETP **67**, 570 (1988)].
[13] A. Hubert, *Theorie der Domanenwände in Geordneten Medien*, Springer-Verlag, Berlin (1974).
[14] A. F. Khapikov, Fiz. Tverd. Tela **36**, 2062 (1994) [Phys. Solid State **36**, 1226 (1994)].
[15] N. L. Schryer and L. R. Walker, J. Appl. Phys. **45**, 5406 (1974).

Translated by D. H. McNeill

Nd_2CuO_4: Chirality and its effect on optical and acoustic properties

E. A. Turov

Institute of Metals, Urals Branch of the Russian Academy of Sciences, 620219 Ekaterinburg, Russia
(Submitted 14 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 1386–1392 (April 1999)

It is shown that two possible magnetic structures in the exchange doublet of the exchange-noncollinear antiferromagnetic material Nd_2CuO_4 that are distinguished by their chirality have certain differences in their optical and acoustic properties. These differences make it possible to identify these structures experimentally. © *1999 American Institute of Physics.* [S1063-7761(99)01604-2]

Among the cuprates, which were the basis for the discovery of widely studied HTSC compounds,[1] one of the most interesting is Nd_2CuO_4. Indeed, it displays noncollinear, rectangular-crosswise magnetic exchange structure, with a number of magnetic and other properties associated with ferromagnetism. At $T \leqslant T_N \simeq 246$ K the crystal chemical symmetry of this cuprate is described by the D_{4h}^{14} $(P4_2/mnm)$ group, and the magnetic Cu^{2+} ions occupy a fourfold $4f$ position. Figure 1a shows the projection of the crystallographic unit cell on the (001) plane, with the cell being chosen in two ways. It is denoted by dashed lines in accordance with Ref. 2, and solid lines in accordance with Refs. 3 and 4 when the coordinate axis is shifted along the x axis by $-1/2$. The copper ions are shown, while the neodymium magnetic moments, which are ordered only at liquid helium temperatures, are not discussed further. The positions of several elements of symmetry are also indicated. The magnetic cell coincides with the crystal chemical cell. Two different magnetic exchange structures (I and II) are shown to the right in Fig. 1b. They have the same exchange energy (the ''exchange doublet''), and correspond to two different orientations of the states (o_1 and o_2), whose energies differ only as a result of relativistic (magnetically anisotropic) interactions.

That these magnetic exchange structures in the magnetic doublet are different is also indicated by the fact that, independent of their orientational states, they have different magnetic symmetries.[3] In fact, for example, the $\mathbf{4}_2$ symmetry axis is retained for structure I of the magnetic group, while it is converted into the ''dashed'' $\mathbf{4}_2' \equiv \mathbf{4}_2 \cdot 1'$ axis (where $1'$ is the time reversal operator) for structure II.

To describe the magnetic exchange structure and properties of Nd_2CuO_4, it is convenient introduce the ferromagnetism and antiferromagnetism basis vectors $\mathbf{M}$ and $\mathbf{L}_n$ $(n = a,b,c)$ instead of the four sublattice magnetizations $\mathbf{M}_\nu$ $(\nu = 1,2,3,4)$:

$$\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \mathbf{M}_4, \tag{1}$$

$$\mathbf{L}_a = \mathbf{M}_1 + \mathbf{M}_2 - \mathbf{M}_3 - \mathbf{M}_4, \tag{2}$$

$$\mathbf{L}_b = \mathbf{M}_1 - \mathbf{M}_2 + \mathbf{M}_3 - \mathbf{M}_4, \tag{3}$$

and

$$\mathbf{L}_c = \mathbf{M}_1 - \mathbf{M}_2 - \mathbf{M}_3 + \mathbf{M}_4. \tag{4}$$

It is important for the following discussion to know how these vectors transform under the symmetry operations of the D_{4h}^{14} group. Because the chemical and magnetic unit cells coincide, translations can be regarded as the symmetry identity element and we may consider only rotations and reflections corresponding to the generators of this group. For the latter it is convenient to take the $\mathbf{4}_2 \| z$ helical symmetry axis, the $\mathbf{2}_{xy} \| [110]$ diagonal binary axis, and the center of symmetry $\overline{1}$ lying at the points of intersection of $\mathbf{4}_2$ with the planes $z = 0$ and $z = 1/2$, where the copper ions are located, as well as at other points separated from these by half the lattice period in any direction. Above all, it is necessary to find those permutations of the ion numbers that produce these symmetry elements, and then easily construct Table I, which lists additional transformations of the vectors (1)–(4) owing to these permutations.

For purposes of precision, it is desirable to elaborate upon the significance of this table. It shows what additional actions on the vectors $\mathbf{L}_a$, $\mathbf{L}_b$, and $\mathbf{L}_c$ produce the elements $\overline{1}$, $\mathbf{4}_2$, and $\mathbf{2}_{xy}$ as elements of the space group as a result of permutations of the atoms beyond the action of the point group. It is evident from the table that this permutation can lead to permutation of the indices a, b, and c, as well as an additional change in sign of several components of the vectors L_n (the permutations have no effect on $\mathbf{M}$). In determining the resulting transformation of the components of these vectors, it is necessary, of course, to include both the ''point effect'' of the symmetry elements (rotations and reflections) and their additional permutational effect, which is illustrated schematically in the table. As an example: $\mathbf{4}_2 L_{by} = L_{ax}$ (instead of $\mathbf{4} L_y = -L_x$ for the point permutation of $\mathbf{4}$).

It follows from Table I that the vector $\mathbf{L}_c$ in (4) only transforms into itself and therefore forms a one-dimensional vector representation corresponding to a collinear magnetic exchange structure with $\mathbf{M}_1 \uparrow\uparrow \mathbf{M}_4 \uparrow\downarrow \mathbf{M}_2 \uparrow\uparrow \mathbf{M}_3$. Matters are different with $\mathbf{L}_a$ and $\mathbf{L}_b$, which transform into one another (for the operations $\mathbf{4}_2$ and $\mathbf{2}_{xy}$), to form a two-dimensional vector representation. These correspond to the noncollinear magnetic exchange structures illustrated in Fig. 1(b).

To distinguish between structures I and II in the exchange doublet, the so-called chirality vector is introduced:[5]

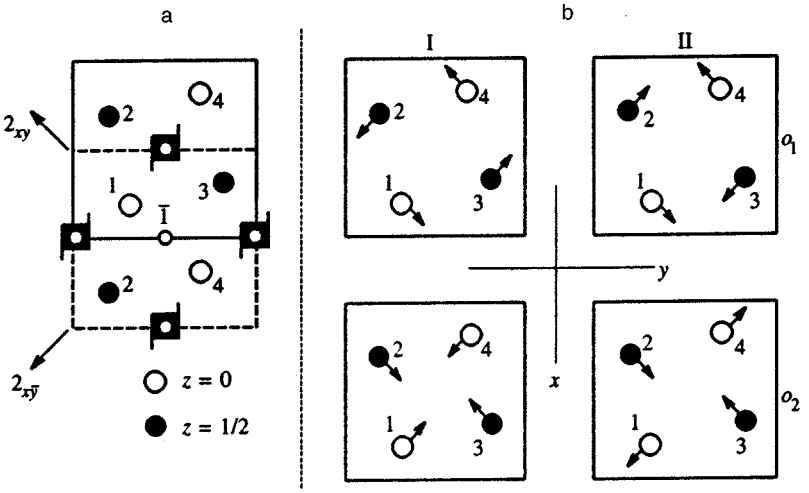

FIG. 1. Projection of the unit cell of Nd_2CuO_4 (group $D_{4h}^{14} \equiv P4_2/mnm$). a) Two choices: with center $\bar{1}$ (dashed lines) or with center 4_2 (smooth lines). The $4f$ sites occupied by Cu^{2+} ions are shown, along with the locations of the symmetry elements: inversion $\bar{1}$, fourth-order helical axis 4_2, and binary axes 2_{xy} and $2_{x\bar{y}}$. b) Possible magnetic structures for chirality $Q_z=+1$ (I) and $Q=-1$ (II) and different orientational states o_1 and o_2, which differ by a $\pi/2$ rotation about the 4_2 axis.

$$\mathbf{Q}=(2M_0)^{-2}([\mathbf{M}_1\mathbf{M}_2]+[\mathbf{M}_2\mathbf{M}_3]+[\mathbf{M}_3\mathbf{M}_4]+[\mathbf{M}_4\mathbf{M}_1])$$
$$=(1/2)(2M_0)^{-2}[\mathbf{L}_a\mathbf{L}_b], \tag{5}$$

where $\mathbf{M}_\nu^2=M_0^2$. With this condition (equal magnitudes), it is easy to find that $Q_x=Q_y=0$ for the magnetic exchange structures I and II shown in the figure, with $Q_z=+1$ for structure I and $Q_z=-1$ for structure II. This holds for both orientational states o_1 and o_2. Therefore, the magnetic exchange states I and II differ in the sign of their chirality, and in the following we refer to them as the $Q(+1)$ and $Q(-1)$ structures.

It is then easy to show, with the aid of the above table of permutation transformations, that the projection Q_z is an invariant of the D_{4h}^{14} group, i.e., of the transformations $\bar{1}$, $\mathbf{4}_2$, and $\mathbf{2}_{xy}$. Since this table only includes the permutation produced by these symmetry elements, in accordance with the above remarks their ''point effect'' must be taken into account separately. Here, for example, the sign change in Eq. (5) associated with the permutation of a and b for $\mathbf{2}_{xy}$ is cancelled by a sign change caused by this element for the z-projection of the vector $\mathbf{Q}$. In other words,

$$\bar{1}Q_z=Q_z, \quad \mathbf{4}_2Q_z=Q_z, \quad 2_{xy}Q_z=Q_z. \tag{6}$$

Proceeding to the optical properties, we write out the invariant expressions for those components of the dielectric permittivity tensor ε_{ij} that depend on Q_z, which are needed to discuss wave vectors $\mathbf{k}\|z$:[6]

$$\varepsilon_{xx}=\varepsilon_\perp+\alpha Q_z+\gamma_1 L_a^2+\gamma_2 L_b^2,$$
$$\varepsilon_{yy}=\varepsilon_\perp+\alpha Q_z+\gamma_1 L_b^2+\gamma_2 L_a^2, \tag{7}$$

and

$$\varepsilon_{xy}=\varepsilon_{yx}^*=i\beta Q_z H_z. \tag{8}$$

TABLE I. Additional transformations of the vectors (1)–(4).

	M	$\mathbf{L}_a$	$\mathbf{L}_b$	$\mathbf{L}_c$
$\bar{1}$	**M**	$-\mathbf{L}_a$	$-\mathbf{L}_b$	$\mathbf{L}_c$
4_2	**M**	$\mathbf{L}_b$	$-\mathbf{L}_a$	$-\mathbf{L}_c$
2_{xy}	**M**	$\mathbf{L}_b$	$\mathbf{L}_a$	$\mathbf{L}_c$

Here ε_{xx} and ε_{yy} include exchange terms[7] proportional to L_a^2 and L_b^2, besides a term with Q_z. But for the equilibrium magnetic exchange structures shown in the figure, $L_a^2=L_b^2$ ($=8M_0^2$ in the equal magnitude model), so that these terms yield the same renormalization of the constant $\varepsilon_\perp$. Retaining the previous notation for the renormalization constant, instead of Eq. (7) we can write

$$\varepsilon_{xx}=\varepsilon_{yy}=\varepsilon_\perp+\alpha Q_z\equiv\varepsilon_Q. \tag{9}$$

Note that all three constants in Eqs. (8) and (9) ($\varepsilon_\perp$, α, and β) generally depend on the frequency ω. Furthermore, β vanishes as $\omega\to 0$.

In ε_{xy} it might be possible to include a symmetric exchange term $\text{const}\cdot\mathbf{L}_a\mathbf{L}_b$.[7] However, for the structures shown in the figure (and by virtue of the equal magnitude condition), for which

$$\mathbf{L}_c=\mathbf{L}_a\mathbf{L}_b=0, \tag{10}$$

this term goes to zero. This also applies to the relativistic terms $L_{ax}L_{bx}+L_{ay}L_{by}=-L_{az}L_{bz}$, which are isotropic in the xy plane.[1)]

Maxwell's equations for the wave fields with $\mathbf{k}\|z$ must therefore include the tensor ε_{ij} defined by Eqs. (8) and (9). The results are circularly polarized waves with refractive indices

$$n_{1,2}^2=\varepsilon_Q\pm|\varepsilon_{xy}| \tag{11}$$

and amplitude ratios (polarizations)

$$\left(\frac{E_x}{E_y}\right)_1=\left(\frac{E_y}{E_x}\right)_2=\frac{\varepsilon_{xy}}{|\varepsilon_{xy}|}=iQ_z\frac{\beta}{|\beta|}. \tag{12}$$

For $H_z=0$,

$$n_1^2=n_2^2=\varepsilon_\perp+\alpha Q_z\equiv n_Q^2, \tag{13}$$

while there is a degeneracy in the polarization, i.e., the phase velocity $v_Q=c/n_Q$ is the same for both modes and does not depend on the direction of **E** in the xy plane. But it does depend on the sign of Q_z. For the two domains with different chirality ($Q_z=+1$ or -1), the velocities are different, with

$$v_{\pm 1}=\frac{c}{\sqrt{\varepsilon_\perp}}\left(1\mp\frac{\alpha}{2\varepsilon_\perp}\right). \tag{14}$$

When $H_z\neq 0$, according to Eqs. (11) and (12) a field of this sort leads to the Faraday effect, whose sign also depends on the sign of Q_z. According to Eq. (12), when the sign of Q_z changes, a right circularly polarized wave becomes left circularly polarized, and *vice versa*. Here the Faraday rotation angle ψ for a linearly polarized wave (per unit length) will be given by

$$\psi=\pm\frac{\omega|\beta H_z|}{2cn_Q}, \tag{15}$$

where the sign is the same as that of the product $\beta Q_z H_z$.

It is scarcely necessary (being so obvious) to explain how this all changes if we include in ε_{xy} (8), and therefore on the right-hand side of Eq. (12), a diamagnetic term proportional simply to H_z (and thus independent of Q_z). We have assumed above that this is small compared to the antiferromagnetic (chiral) contribution.

Thus far we have been dealing with ''domain chirality,'' having in mind that $Q(+1)$ and $Q(-1)$ domains can coexist, as they have the same exchange energies (the exchange doublet). In this case, the effects considered above might make it possible to observe a chiral domain structure of this type.

In fact, however, it has been found experimentally (see Scanthakumar *et al.*[8] and references therein) and explained theoretically by Blinkin *et al.*[3] and Vitebskii *et al.*[4] that the exchange degeneracy in the $Q(+1)$ and $Q(-1)$ doublet is removed by a magnetic anisotropy, so that these structures develop over different temperature ranges. As the temperature decreases following the transition into the antiferromagnetic region at a temperature of $T_{N1}=246$ K, the magnetic copper ions form the $Q(+1)$ magnetic exchange structure Io_1 in the figure. This is referred to as phase I in Ref. 8 and phase τ_2 in Refs. 3 and 4. Then at $T_{N2}=75$ K there is a sudden transition to a $Q(-1)$ magnetic exchange structure (structure IIo_1 in the figure), which is referred to as phase II in Ref. 8 or phase τ_8 in Refs. 3 and 4.

As can be seen from the figure, the transition between the $Q(+1)$ and $Q(-1)$ phases can proceed, for example, through a simultaneous reversal of the magnetic moments 2 and 3 (i.e., in the $z=1/2$ layer) while moments 1 and 4 are unchanged (i.e., in the $z=0$ layer). As a result, there is a phase transition from a phase with chirality $Q=(+1)$ to one with $Q=(-1)$ (or *vice versa*). This is probably a first-order phase transition, since a change in chirality through a continuous in-phase rotation of the magnetic moments is impossible: the $Q(+1)$ and $Q(-1)$ phases are topologically distinct. Finally, at $T=T_{N3}=30$ K there is a reverse phase transition $Q(-1)\rightarrow Q(+1)$. Subsequently, at liquid helium temperatures, where the magnetic Nd ions already play a significant role, the magnetic structure of the copper ions becomes collinear (and for $T=1.5$ K the magnetic moments of the Nd ions become ordered).

It might be hoped that the phase transitions described above, with a $Q(+1)\leftrightarrow Q(-1)$ change in chirality, would show up as a change in the optical properties determined by Eqs. (11)–(15). For $H_z=0$ the velocity of light should change at the phase transition point, and for $H_z\neq 0$ the sign of the Faraday rotation angle should change as well.

Analogous acoustic effects associated with chirality probably also exist. In any case, this is of importance for chiral antiferromagnetic materials, which are not optically transparent.

Let us again consider an elastic wave with $\mathbf{k}\|z$. With Eq. (6) and the table it is easy to show that the invariant expressions for the elastic moduli C_{ijkl} that determine this sound can be written in the form

$$C_{xzxz}=C_{yzyz}=C_{44}+\alpha Q_z\equiv C_Q, \tag{16}$$

and

$$C_{xzyz}=-C_{yzxz}=i\beta Q_z H_z. \tag{17}$$

Here exchange (isotropic) terms of the type L_a^2 and L_b^2 are introduced through C_{44}, as for ε_{xx} and ε_{yy} [cf. Eq. (7)].

The solutions of the dynamic equations for elastic systems, allowing for Eqs. (16) and (17), are circularly polarized transverse waves whose wave numbers and polarizations are given by

$$k_{1,2}\simeq\frac{\omega}{v_Q}\left(1\mp\frac{|\beta H_z|}{2C_{44}}\right),\quad v_Q=(C_Q/\rho)^{1/2}, \tag{18}$$

and

$$\left(\frac{u_x}{u_y}\right)_1=\left(\frac{u_y}{u_x}\right)_2=iQ_z\frac{\beta}{|\beta|}. \tag{19}$$

In the absence of a field ($H_z=0$), the normal modes are degenerate with respect to the wave numbers and polarizations, so that their speed $v_1=v_2\equiv v_Q$ and is independent of the direction of the elastic displacements $\mathbf{u}$ in the xy plane. But this velocity depends on the chirality Q_z, since it is different for the $Q(+1)$ and $Q(-1)$ phases.

When $H_z\neq 0$, this degeneracy is removed, so that even for the same chirality Q_z the velocities $v_{1,2}=\omega/k_{1,2}$ are different according to Eq. (13). In addition, according to Eq. (19), the sign of the acoustic Faraday rotation depends on the sign of Q_z, which changes during a $Q(+1)\leftrightarrow Q(-1)$ phase transition.

In conclusion, we briefly recall the possible origin of the energy difference between the $Q(+1)$ and $Q(-1)$ phases. According to Eq. (6), the point is that the thermodynamic potential contains a term of the form

$$\text{const}\cdot Q_z, \tag{20}$$

which, in particular, can also produce this energy difference. As it is isotropic in the xy plane [according to Eq. (20)], this term is highly reminiscent of antisymmetric Dzyaloshinskiĭ–Moriya exchange, and like the latter, is probably a semirelativistic semi-exchange term. The term (20) of the thermodynamic potential is contained implicitly in the equations of Refs. 3 and 4. It can be obtained from the magnetic anisotropy energy if terms isotropic in the xy basis plane are separated out.

As noted above, this discussion of certain features of the optical and acoustic properties of Nd_2CuO_4 is based on the results of Refs. 3 and 4, which provide a fairly convincing (from symmetry considerations) description of the Io_1 and IIo_1 magnetic structures (see Fig. 1) and the phase transitions between them observed in neutron diffraction measurements.[8] However, there are a great many other experiments devoted to magnetic structure and phase transitions in Nd_2CuO_4, including some that appeared later. These do not always agree with one another or with Ref. 8. The discussion seems to have been summarized in an article[9] that actually confirms the results of Ref. 8, and therefore the existence of type I and II structures with differing chiral structures and phase transitions between them. (To avoid possible confusion in comparing our Fig. 1 with the corresponding figures in Ref. 9, note that the latter correspond to a crystal structure that exists at temperatures above the structural transition point, $T_c \approx 300$ K.)

These remarks again confirm the appropriateness of introducing chirality, which makes it possible to identify chiral magnetic structures by means of optical and acoustic measurements.

We note, finally, that the above effects of chirality on the properties of Nd_2CuO_4 should also show up in other cuprates of the form R_2CuO_4 (in particular, when R denotes Pr or Sm), which have analogous noncollinear magnetic exchange structure.[9]

The author thanks M. I. Kurkin, V. V. Men'shenin, V. E. Naĭsh, and V. V. Ustinov for useful discussions and valuable comments. This work was supported by the Russian Fund for Fundamental Research (Grant No. 96-02-16489).

[1)]The author is sincerely grateful to Yu. G. Pashkevich, who kindly furnished a clarification of Ref. 7.

[1] W. E. Picket, Rev. Mod. Phys. **61**, 433 (1989).
[2] *International Tables for Crystallography*, Birmingham (1952, 1965, 1983).
[3] V. A. Blinkin, I. M. Vitebskiĭ, O. D. Kolotiĭ *et al.*, Zh. Éksp. Teor. Fiz. **98**, 2098 (1990) [Sov. Phys. JETP **71**, 1179 (1990)].
[4] I. M. Vitebskiĭ, N. M. Lavrinenko, and V. L. Sobolev, J. Magn. Magn. Mater. **97**, 263 (1990).
[5] H. Kawamura and S. Miyashima, J. Phys. Soc. Jpn. **53**, 4138 (1984).
[6] E. A. Turov, *Kinetic, Optical, and Acoustic Properties of Antiferromagnetic Materials* [in Russian], Uro AN SSSR (Urals Branch, Russian Academy of Sciences), Sverdlovsk (1990).
[7] I. M. Vitebskiĭ, A. V. Yeremenko, Yu. G. Pashkevich *et al.*, Physica C **178**, 189 (1991).
[8] S. Scanthakumar, H. Zhang, T. W. Clinton *et al.*, Physica C **160**, 124 (1989).
[9] R. Sachidanandum, T. Yildirum, A. B. Harris *et al.*, Phys. Rev. B **56**, 260 (1997).

Translated by D. H. McNeill

Ferromagnetism of binary compounds with cubic symmetry

R. O. Zaĭtsev*)

Kurchatov Institute, 123182 Moscow, Russia

Yu. V. Mikhaĭlova

''NIIteplopribor'' State Scientific Center, 129085 Moscow, Russia

(Submitted 5 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1393–1410 (April 1999)

Based on the idea of a strong interaction within the same unit cell, the possible existence of a ferromagnetic instability in a system with jumps from transition element cations to non-transition element anions and vice versa is established. A phase diagram is constructed for the ferromagnetic ordering as a function of the degree of filling, n_p and n_d, of the p^6- and d^{10}-shells of non-transition and transition elements, respectively. © *1999 American Institute of Physics.* [S1063-7761(99)01704-7]

1. STATEMENT OF THE PROBLEM AND GENERAL EQUATIONS

In this paper we study the electronic properties of the simplest metallic binary compounds with a CsCl-type structure (see Fig. 1 below).

The magnetic properties are studied in terms of a generalized Hubbard model in which the isolated atom states determined by Hund's rule are used as a zeroth approximation.

The interaction energy of the electrons within a given atom (the so-called Hubbard energy) is assumed to be the largest energy parameter and is assumed to be infinite from the start.

In the metallic phase, the long range part of the Coulomb and exchange interactions are substantially screened, so that in the following we only include the interaction between nearest neighbors owing to overlap of their wave functions:

$$\hat{H}=\sum_{\mathbf{r},\mathbf{r}'} V^{\alpha,k}(\mathbf{r},\mathbf{r}')\{\hat{e}^{+}_{\alpha,\sigma}(\mathbf{r})\hat{p}_{k,\sigma}(\mathbf{r}')+\text{H.c.}\}$$
$$+\sum_{\mathbf{r},\sigma}(\epsilon_e-\sigma H)\hat{e}^{+}_{\alpha,\sigma}(\mathbf{r})\hat{e}_{\alpha,\sigma}(\mathbf{r})$$
$$+\sum_{\mathbf{r},\sigma}(\epsilon_p-\sigma H)\hat{p}^{+}_{k,\sigma}(\mathbf{r})\hat{p}_{k,\sigma}(\mathbf{r}). \quad (1)$$

Here the indices α take two values corresponding to the two degenerate states, $\sqrt{3}(x^2-y^2)$ and $3z^2-r^2$, of the e_g-electrons of a transition cation, and the indices k take three values corresponding to the three degenerate states, $a=p_x$, $b=p_y$, and $c=p_z$, of the p-electrons of a non-transition anion.

The energy difference $r=\epsilon_p-\epsilon_e$ is assumed to be a given parameter, while the sum of the one-particle energies can be expressed in terms of the chemical potential $\mu=-(\epsilon_p-\epsilon_e)/2$.

After transforming to a Fourier representation, the overlap integral matrix $\hat{V}^{\alpha,k}(\mathbf{q})$ is easily calculated for a given crystal structure.[1]

For a given location of the lowest states of the atomic multiplets, the collectivization of the transitions between the N- and $(N+1)$-states is determined by the poles of the one-particle Green function. To calculate it, we write the expansion of the creation and annihilation operators for all possible transitions between states of the lowest multiplets:

$$\hat{e}^{+}_{\alpha,\sigma}(\mathbf{r})=\sum_{\beta} g_\beta(\alpha,\sigma)\hat{X}^{\beta}_{\mathbf{r}}, \quad \hat{p}_{k,\sigma}(\mathbf{r})=\sum_{v} g_v(k,\sigma)\hat{Y}^{v}_{\mathbf{r}}. \quad (2)$$

Here β and v are the numbers of the one-particle e- or p-transitions and g_β and g_v are the so-called genealogical coefficients, which we find for each specific type of transition.

In the end, it is necessary to calculate the average occupation number of each component of the multiplet as a function of the magnitude of the applied magnetic field H.

In the following we use the zero-loop approximation or Hubbard I approximation,[2] in which each self-energy part of the one-particle Green function is replaced by the so-called end factor f_β or f_v, which equals the sum of the occupation numbers of the initial and final states responsible for a given transition β or v.

In this approximation[3] the complete Green function $D_\omega(\mathbf{q})$ is the product of the virtual Green function $G^{\alpha,v}_{\omega}(\mathbf{q})$ and the end factor f_v. The virtual Green function itself satisfies a Dyson-type equation:

$$D^{\alpha,v}_{\omega}(\mathbf{q})=G^{\alpha,v}_{\omega}(\mathbf{q})f(\mathbf{v}), \quad \hat{G}^{-1}_{\omega}(\mathbf{q})=(\hat{G}^{(0)}_{\omega})^{-1}-\hat{W}. \quad (3)$$

The elements of the matrix $\hat{W}$ are determined from the coefficients in the expansion (2) for the creation and annihilation operators after they are substituted in the Hamiltonian (1):

$$W_{\beta,v}(\mathbf{q})=f_{(\beta)}\sum_{\alpha,k} g_\beta(\alpha)\sum_{k} V^{\alpha,k}(\mathbf{q})g_v(k). \quad (4)$$

All possible $N+1$-particle occupation numbers corresponding to the transition $\alpha(N_k,(N+1)_s)$ at a specified tempera-

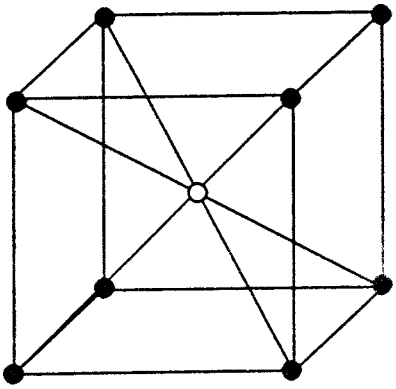

FIG. 1. CsCl-type cubic lattice.

ture T and chemical potential are calculated from the diagonal matrix elements of the one-particle Green function with the aid of the following general formula:

$$n_{(N+1)}(s)=T\sum_{\omega,\mathbf{q}}\exp(i\omega\delta)D_{\omega}^{\alpha,-\alpha}(\mathbf{q}). \tag{5}$$

Here δ is a positive infinitesimal quantity, $\omega=(2n+1)\pi T$, and the symbol $-\alpha$ labels the transition that is the inverse of α.

The right-hand side of Eq. (5) can be found using Eqs. (3) and (4) in terms of the end factors that appear in the matrix elements for the Green function and its inverse.

The average occupation numbers can be expressed in terms of the end factors, so that in the approximation used here it is possible to obtain a closed system of equations for the variations in all the end factors as a function of the applied magnetic field.

Ultimately, it is possible to obtain an equation for the magnetic permeability as a function of temperature and the average number of n_e- and n_p-electrons belonging to a single unit cell.

The ferromagnetic instability is defined by an infinite spin magnetic susceptibility.

The method proposed here explains the ferromagnetism of compounds of type FeAl, with a Curie temperature $T_c=623\,\mathrm{K}$ and $n^*\approx 1\mu_B$, and MnSi with a Curie temperature $T_c=34\,\mathrm{K}$ and $n^*\approx 0.4\mu_B$. Assuming that the t_{2g}-shell of the transition element and the $3s$-shell of the nontransition element are completely filled, then the total number $n_e+n_p=3-n_s$, where n_s is the average number of $4s$-electrons of the transition element, which we take to be an arbitrary parameter.

For this reason, in the following we examine the simultaneous filling of the e_g- and $3p$-shells in detail for all integer ranges of n_p and n_e.

The matrix of the transition integrals between nearest ions is proportional to the overlap integral $(16/\sqrt{3})\tau(111)_{x,x^2-y^2}$, which is assumed equal to unity below. The remaining matrix elements are calculated according to the *fcc* unit cell structure shown in Fig. 1:

$$\hat{V}^{\alpha,k}(\mathbf{q})=\begin{matrix}\sqrt{3}(x^2-y^2)\\ 3z^2-r^2\end{matrix}$$
$$\times\begin{matrix}p_x\\ \\ \end{matrix}\begin{pmatrix}\sqrt{3}\nu_x^-\nu_y^+\nu_z^+ & -\sqrt{3}\nu_x^+\nu_y^-\nu_z^+ & 0\\ -\nu_x^-\nu_y^+\nu_z^+ & -\nu_x^+\nu_y^-\nu_z^+ & 2\nu_x^+\nu_y^+\nu_z^-\end{pmatrix}\begin{matrix}\\ p_y\\ \\ \end{matrix}\begin{matrix}\\ p_z\\ \\ \end{matrix} \tag{6}$$

where the matrix elements depend on the quasimomenta $q_{x,y,z}$ through the functions $v_k^{\pm}=1\pm\exp(iq_k)$.

2. CALCULATING THE PHASE DIAGRAM

Low concentrations, $n_{p,e}<1$. In the limiting case of infinite Hubbard energy, the creation and annihilation operators can be conveniently expressed in terms of $\hat{X}$, the transition operators between vacant $|0\rangle$ and one-particle $|\sigma,\lambda\rangle$ states:

$$\hat{e}_{\mathbf{r}\sigma,\eta}=\hat{X}_{\mathbf{r}}^{(0|\sigma,\eta)},\quad \hat{e}_{\mathbf{r}\sigma,\eta}^{+}=\hat{X}_{\mathbf{r}}^{(\sigma,\eta|0)};$$
$$\hat{p}_{\mathbf{r}\sigma,\nu}=\hat{Y}_{\mathbf{r}}^{(0|\sigma,\nu)},\quad \hat{p}_{\mathbf{r}\sigma,\nu}^{+}=\hat{Y}_{\mathbf{r}}^{(\sigma,\nu|0)}. \tag{7}$$

After this expansion is substituted in the initial Hamiltonian (1) we obtain an explicit expression for the inverse Green function:

$$\hat{G}_{\omega}^{-1}(\mathbf{q})=\begin{matrix}\\ e_\alpha\\ \\ p_k\end{matrix}\begin{matrix}\begin{matrix}e_\beta & \quad p_s\end{matrix}\\ \begin{pmatrix}(i\omega-\epsilon_e)\delta_{\alpha,\beta} & -f_\alpha V^{\alpha,k}(\mathbf{q})\\ -f_p V^{k,\beta}(\mathbf{q}) & (i\omega-\epsilon_p)\delta_{k,s}\end{pmatrix}\end{matrix}. \tag{8}$$

Here the $\epsilon_{p,e}$ are the energies of the one-particle p- and e-states, $f_{p,e}$ are the end factors given by the sum of the occupation numbers of the initial and final states, $V^{\alpha,s}(\mathbf{q})$ is the energy matrix (6) for the transition, and $V^{k,\beta}(\mathbf{q})$ is the Hermitian conjugate matrix.

To find the average occupation numbers $n_{p,e}^{(\sigma)}$ we use the simplest approximation,[2] in which the excitation energy is determined by the average self-energy part, which can be expressed in terms of the so-called end factors, which, in this case, equal the sum of the average number of vacant and one-particle states,

$$f_{p,k}^{(\sigma)}=n_0+n_{I,k}^{(\sigma)},\quad f_{e,\alpha}^{(\sigma)}=n_0+n_{I,\alpha}^{(\sigma)}. \tag{9}$$

Given that these are independent of the number k of the one-particle state and α, we obtain the following equations of state:

$$n_e^{(\sigma)}=\sum_{\alpha}n_{I,\alpha}^{(\sigma)}=2f_e^{(\sigma)}\sum_{\mathbf{p}\lambda=\pm}a_{\mathbf{p}}^{(-\lambda)}n_F(\xi_{\mathbf{p}}^{(\sigma,\lambda)}), \tag{10}$$

$$n_p^{(\sigma)}=\sum_{k}n_{I,k}^{(\sigma)}=f_p^{(\sigma)}\left[n_F(\epsilon_p)+2\sum_{\mathbf{p}\lambda=\pm}a_{\mathbf{p}}^{(\lambda)}n_F(\xi_{\mathbf{p}}^{(\sigma,\lambda)})\right]. \tag{10$'$}$$

Here $n_F(\epsilon)$ is the Fermi distribution and the two doubly degenerate branches of the spectrum are

$$\xi_{\mathbf{p}}^{(\sigma,\pm)}=\pm\sqrt{\left(\frac{r}{2}\right)^2+f_p^{(\sigma)}f_e^{(\sigma)}t_{\mathbf{p}}^2}-\sigma H-\mu,$$

$$a_{\mathbf{p}}^{(\pm)}=\frac{1}{2}\left[1\pm\frac{\tau}{\xi_{\mathbf{p}}^{(\sigma,+)}-\xi_{\mathbf{p}}^{(\sigma,-)}}\right], \tag{11}$$

where

$$t_{\mathbf{p}}^2=s_x^2c_y^2c_z^2+s_y^2c_z^2c_x^2+s_z^2c_x^2c_y^2, \quad s_k^2=4\sin^2(p_k/2),$$

$$c_k^2=4\cos^2(p_k/2).$$

The equations of state (10) can be expressed in terms of integrals of the single variable t_p^2, so it makes sense to introduce the density of states function

$$\rho(\epsilon)=\sum_{p_x,p_y,p_z}\delta(\epsilon-t_{\mathbf{p}}^2),$$

which can be used to write all the sums over the momentum.

Given the cubic symmetry, the end factors (9) can be expressed in terms of the average occupation numbers of the one-particle states,

$$n_{p,e}^{(\sigma)}=\sum_{s=1}^{3,2}n_s^{(\sigma)}.$$

Since

$$n_0+\sum_{s,\sigma}n_s^{(\sigma)}=1,$$

we obtain

$$f_p^{(\sigma)}=1-n_p^{(-\sigma)}-\frac{2}{3}n_p^{(\sigma)}, \quad f_d^{(\sigma)}=1-n_e^{(-\sigma)}-\frac{1}{2}n_e^{(\sigma)}. \tag{12}$$

From this we find the variation in the end factors in terms of the variation in the average occupation numbers, which are determined by the variation in the external magnetic field:

$$\delta n_{p,e}^{(\sigma)}=-\delta n_{p,e}^{-(\sigma)}, \quad \delta f_p^{(\sigma)}=\frac{1}{3}\delta n_p^{(\sigma)}, \quad f_e^{(\sigma)}=\frac{1}{2}\delta n_e^{(\sigma)}. \tag{13}$$

These formulas, along with the equations of state (10) for a given magnetic field, can be used to write an equation for just the variation in the end factors:

$$\delta f_e^{(\sigma)}=[K_e+L_e]\delta f_e^{(\sigma)}+\frac{f_e}{f_p}L_d\delta f_p-f_eR_e\sigma\delta H,$$

$$\delta f_p^{(\sigma)}=[K_p+L_p]\delta f_p^{(\sigma)}+\frac{f_p}{f_e}L_p\delta f_e-f_pR_p\sigma\delta H. \tag{14}$$

This yields the magnetic susceptibility

$$\chi=\frac{\delta n_e^\sigma}{\delta H}+\frac{\delta n_p^\sigma}{\delta H}=2\frac{\delta f_e^\sigma}{\delta H}+3\frac{\delta f_p^\sigma}{\delta H}. \tag{15}$$

We obtain the possibility of a ferromagnetic instability as the condition for a singularity in the magnetic susceptibility:

$$\det\begin{pmatrix}1-(K_e+L_e), & -f_eL_e/f_p\\ -f_pL_p/f_e, & 1-(K_p+L_p)\end{pmatrix}=0. \tag{16}$$

Here we have introduced the functions

$$K_e=\sum_{\mathbf{p},\lambda=\pm}a_{\mathbf{p}}^{(-\lambda)}n_F(\xi_{\mathbf{p}}^{(\lambda)})=\frac{n_e}{4-3n_e},$$

$$\xi_{\mathbf{p}}^{(\pm)}=\pm\sqrt{(r/2)^2+f_ef_pt_{\mathbf{p}}^2}-\mu, \tag{17}$$

and

$$K_p=\frac{1}{3}\left[n_F(\epsilon_p)+2\sum_{\mathbf{p},\lambda=\pm}a_{\mathbf{p}}^{(\lambda)}n_F(\xi_{\mathbf{p}}^{(\lambda)})\right]=\frac{n_p}{6-5n_p}. \tag{18}$$

These functions all depend only on $f_pf_et_p^2$, which is invariant under the cubic symmetry transformation [defined in Eq. (5)].

The coefficients $L_{p,e}$ in the equation that determine the limits for ferromagnetic ordering are defined in terms of the derivatives of the right-hand side of the equations of state:

$$L_{p,e}=b\frac{\partial K_{p,e}}{\partial b}, \quad \text{where } b=g_p^2g_e^2f_pf_et^2. \tag{19}$$

In this case the coefficients $g_{p,e}^2=1$ and

$$f_e=\frac{4-3n_e}{4}, \quad f_p=\frac{6-5n_p}{6}. \tag{20}$$

In this range, the electronic states are resonant between the vacant and one-particle states. Thus, for a small number of excitations, no ferromagnetic state develops,[3] but the system has elevated magnetic susceptibility.

Concentrations $n_p<1$, $1<n_e<2$. Let us examine the situation in which e_g-electrons resonate between one- and two-particle states, while, as before, the number of p-electrons is less than one ($n_p<1$).

The one-particle states $\hat{e}_{1,\sigma}^+|0\rangle$ and $\hat{e}_{2,\sigma}^+|0\rangle$ have spin 1/2. The lowest energy two-particle states $^3A_2$ have spin $S=1$:

$$\hat{e}_{1,\sigma}^+\hat{e}_{2,\sigma}^+|0\rangle \quad (S_z=\sigma),$$

$$\frac{\hat{e}_{1,\uparrow}^+\hat{e}_{2,\downarrow}^++\hat{e}_{1,\downarrow}^+\hat{e}_{2,\uparrow}^+}{\sqrt{2}}|0\rangle \quad (S_z=0). \tag{21}$$

The higher energy ^1E and $^1A_1$ states are neglected.

For a finite external magnetic field, the variation in the end factors depends on the variation in both the one-particle (n_{I}) and two-particle (n_{II}) occupation numbers. Given the symmetry of the system under interchange of the $e_2=3z^2-r^2$ and $e_1=\sqrt{3}(x^2-y^2)$ states, we find the variation in the end factors to be

$$f_1^{(\sigma)}=n_{\mathrm{II}}^{(\sigma)}+n_{\mathrm{I}}^{(\sigma)}, \quad \delta f_1^{(\sigma)}=\delta n_{\mathrm{II}}^{(\sigma)}+\delta n_{\mathrm{I}}^{(\sigma)},$$

$$f_2^{(\sigma)}=n_{\mathrm{II}}^{(0)}+n_{\mathrm{I}}^{(-\sigma)}, \quad \delta f_2^{(\sigma)}=\delta n_{\mathrm{I}}^{(-\sigma)}=-\delta n_{\mathrm{I}}^{(\sigma)}. \tag{22}$$

Therefore, in contrast to the previous ''one-particle'' case, here it is necessary to have two independent equations for the variations in the one- and two-particle e-states.

In order to obtain these equations, we multiply the real part of the expansion of the annihilation operator,

$$\hat{e}_{\mathbf{r}\sigma}=g_1\hat{X}_{\mathbf{r}}^{(0,\sigma|\sigma,\sigma)}+g_2\hat{X}_{\mathbf{r}}^{(0,-\sigma|\sigma,-\sigma)}, \tag{23}$$

where $g_1=1$ and $g_2=1/\sqrt{2}$, by an arbitrary linear combination of the conjugate X-operators,

$$\hat{Y}_{\mathbf{r}}=\beta_1\hat{X}_{\mathbf{r}}^{(\sigma,\sigma|0,\sigma)}+\beta_2\hat{X}_{\mathbf{r}}^{(\sigma,-\sigma|0,-\sigma)}.$$

Averaging the individual T-products over the states with given temperature and chemical potential, we find an equation relating the two particle occupation numbers $n_{\rm II}$, the Fourier components of the virtual one-particle Green function $\hat{G}_\omega(\mathbf{p})$, and the end factors $f_k(\sigma)$, where $k=1,2$, and $f_p(\sigma)$.

We now calculate the T-products in terms of the zero loop Hubbard I approximation:

$$-\langle \hat{T}(\hat{e}_{\mathbf{r},\sigma}(\tau),\hat{Y}_{\mathbf{r}}(\tau+0))\rangle$$
$$=(g_1\gamma_1)\langle X^{(S_z=\sigma|S_z=\sigma)}\rangle+(g_2\gamma_2)\langle X^{(S_z=0|S_z=0)}\rangle$$
$$=(g_1\gamma_1)n_{\rm II}^{(\sigma)}+(g_2\gamma_2)n_{\rm II}^{(0)}=T\sum_{\omega,\mathbf{p};\alpha,\beta} g_\alpha G_\omega^{(\alpha,\beta)}$$
$$\times(\mathbf{p})\gamma_\beta f_\beta \exp(i\omega\delta). \tag{24}$$

Here $\delta=0+$ is a small positive correction and the f_s are the end factors (22).

The matrix elements of the one-particle Green function can be expressed in terms of the inverse matrix, where now each row and column corresponding to the e-states has twice the number of components. The corresponding matrix elements are given by the genealogical coefficients g_α in the expansion (28). Thus, for the selected operator $\hat{e}_{1\sigma}$, the first two rows of the inverse Green function can be written

$$\hat{G}_\omega^{-1}(\mathbf{p})=\begin{matrix}((0,\sigma)\to(\sigma,\sigma));\\((0,-\sigma)\to(\sigma,-\sigma));\end{matrix}$$

$$\begin{matrix}&\beta=1&\beta=2&s=3,4\\ \alpha=1&\\ \alpha=2&\\ k=3,4...&\end{matrix}\begin{pmatrix} i\omega-\epsilon_e & 0 & -f_1g_1\rho_s^p\\ 0 & i\omega-\epsilon_e & -f_2g_2\rho_s^p\\ -\nu_p^kg_1 & -\nu_p^kg_2 & \tau_{k,s}\end{pmatrix}. \tag{25}$$

Here $g_1=1$, $g_2=1/\sqrt{2}$, and the end factors are given by Eq. (22).

Using the explicit form of the inverse matrix elements, it is possible to calculate the sums on the right-hand side of Eq. (25). Noting that the energies of the transitions accompanying the creation of the same e_1-state are the same, we obtain

$$\sum_{\alpha=1,2} g_\alpha G_\omega^{\alpha,\beta}(\mathbf{p})=g_\beta(i\omega-\epsilon_e)\frac{\det\tau_{k,s}}{\det G^{-1}}, \tag{26}$$

Taking the sum on both sides of this equation over the momenta and frequencies, we find a result proportional to the genealogical coefficients g_β:

$$T\sum_{\omega,\mathbf{p}}\sum_{\alpha=1,2} g_\alpha G_\omega^{\alpha,\beta}(\mathbf{p})=g_\beta K_e(H).$$

As a result, we have an equation for arbitrary γ_α:

$$g_1\gamma_1 n_{\rm II}^{(S_z=\sigma)}+g_2\gamma_2 n_{\rm II}^{(S_z=0)}=K_e(H)\sum_{\beta=1,2} g_\beta\gamma_\beta f_\beta. \tag{27}$$

By varying the magnetic field in Eq. (27), we obtain two equations.

When

$$\sum_{1\leqslant\alpha\leqslant 2} g_\alpha\gamma_\alpha=0$$

we find an equation that is independent of the applied external field,

$$\delta n_{\rm II}^{(\sigma)}(1-K_e)-2K_e\delta n_{\rm I}^{(\sigma)}=\delta f_1^{(\sigma)}(1-K_e)$$
$$+\delta f_2^{(\sigma)}(1+K_e)=0. \tag{28}$$

Here the electron density lies within the interval $1<n_e<2$ and the coefficients are calculated for a vanishing external magnetic field:

$$K_e=\sum_{\mathbf{p},\lambda=\pm} a_{\mathbf{p}}^{(-\lambda)}n_F(\xi_{\mathbf{p}}^{(\lambda)})=2\frac{n_e-1}{2+n_e},$$
$$\xi_{\mathbf{p}}^{(\pm)}=\pm\sqrt{\left(\frac{r}{2}\right)^2+g_e^2f_ef_pt_{\mathbf{p}}^2}-\mu, \tag{29}$$

and

$$f_e=\frac{2+n_e}{12},\quad f_p=\frac{6-5n_p}{6},\quad g_e^2=g_1^2+g_2^2=\frac{3}{2}. \tag{30}$$

When $g_\alpha=\gamma_\alpha$ we obtain an equation for the susceptibilities:

$$\delta n_{\rm II}^{(\sigma)}=\delta f_2^{(\sigma)}+\delta f_1^{(\sigma)}=[K_e+L_e]\sum_{\alpha=1,2} g_\alpha^2\delta f_\alpha^{(\sigma)}$$
$$+g_e^2\frac{f_e}{f_p}L_e\delta f_p-g_e^2f_eR_e\sigma\delta H, \tag{31}$$

where the coefficient L_e is given by the general equation

$$L_e=b\frac{\partial}{\partial b}K_e=\sum_{\mathbf{p},\lambda=\pm} t_{\mathbf{p}}^2\frac{\delta}{\delta t_{\mathbf{p}}^2}[n_F(\xi_{\mathbf{p}}^{(\lambda)})a_{\mathbf{p}}^{(-\lambda)}],$$
$$R_e=\sum_{\mathbf{p},\kappa=\pm} a_{\mathbf{p}}^{(-\kappa)}n_F'(\xi_{\mathbf{p}}^{(\kappa)}),$$
$$a_{\mathbf{p}}^{(\pm)}=\frac{1}{2}\left[1\pm\frac{r}{\sqrt{r^2+4g_e^2f_ef_pt_{\mathbf{p}}^2}}\right]. \tag{32}$$

An equation for the variation in the one-particle p-states can be found from the equation for the occupation numbers n_p in a way fully analogous to Eq. (10):

$$n_p^{(\sigma)}=f_p^{(\sigma)}\left[n_F(\epsilon_p)+2\sum_{\mathbf{p}\kappa=\pm} a_{\mathbf{p}}^{(\kappa)}n_F(\xi_{\mathbf{p}}^{(\sigma,\kappa)})\right]. \tag{33}$$

We find the relationship between the variation in the occupation numbers and that of the end factors $f_p^{(\sigma)}$ using the general relation (14):

$$\delta f_p^{(\sigma)}=L_p\frac{f_p}{g_e^2f_e}\sum_{\alpha=1,2} g_\alpha^2\delta f_\alpha^{(\sigma)}+[K_p+L_p]\delta f_p$$
$$-f_pR_p\sigma\delta H. \tag{34}$$

The coefficients K_p and L_p are calculated using Eqs. (18) and (19) with the excitation energy $\xi_{\mathbf{p}}^{\pm}$ given by Eq. (29).

Therefore, a system of three equations (28), (31), and (34) determines the change in the three end factors, which determine the changes in all three occupation numbers.

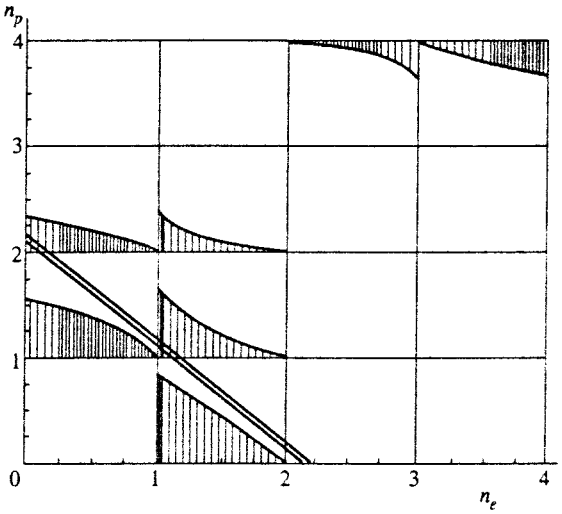

FIG. 2. Magnetic phase diagram at $T=0$. Ferromagnetic regions are shaded. Straight lines correspond to the equations of electrical neutrality, with the upper for MnSi and the lower for FeAl.

The solubility of the corresponding homogeneous system of equations determines whether ferromagnetic instability will develop. As a result, the condition for a phase transition into the ferromagnetic state has the following simple form:

$$(1-K_p)[K_e(1-K_e)-L_e(1/3+K_e)]-L_pK_e(1-K_e)=0. \tag{35}$$

Here

$$K_e=4\frac{(n_e-1)}{2+n_e},\quad K_p=\frac{n_p}{6-5n_p},$$
$$n_e=\frac{2K_e+4}{4-K_e},\quad n_p=\frac{6K_p}{1+5K_p}. \tag{36}$$

This equation establishes a relationship between the energies ϵ_p and ϵ_e. Eliminating them using the equations of state (12) and (17) in zero field, we obtain the magnetic phase diagram in the variables n_e and n_p (see Fig. 2).

Concentrations $n_e<1$, $1<n_p<2$. We now examine the situation in which the states $p_x=a$, $p_y=b$, and $p_z=c$ resonate among one- and two-particle states, while the number of d-electrons in the e_g-shell remains below unity ($n_e<1$).

The one-particle states $\hat{a}^+_\sigma|0\rangle$, $\hat{b}^+_\sigma|0\rangle$, and $\hat{c}^+_\sigma|0\rangle$ have spin 1/2. The lowest energy two-particle states $^3A_2$ have spin $S=1$:

$$\hat{a}^+_\sigma\hat{b}^+_\sigma|0\rangle\quad(S_z=\sigma),\quad\frac{\hat{a}^+_\uparrow\hat{b}^+_\downarrow+\hat{a}^+_\downarrow\hat{b}^+_\uparrow}{\sqrt{2}}|0\rangle\quad(S_z=0). \tag{37}$$

We find the six remaining two-particle states via the cyclic permutation $a\to b\to c\to a$. The higher-energy states ^1E and $^1A_1$ are neglected.

For a finite external magnetic field, the variations in the end factors depend on the variations in both the one ($n_{\rm I}$) and two-particle ($n_{\rm II}$) occupation numbers. Given the symmetry of the system under permutation of the a-, b-, and c-states, we find the variations in the end factors to be

$$f_1^{(\sigma)}=n_{\rm II}^{(\sigma)}+n_{\rm I}^{(\sigma)},\quad \delta f_1^{(\sigma)}=\delta n_{\rm II}^{(\sigma)}+\delta n_{\rm I}^{(\sigma)},$$
$$f_2^{(\sigma)}=n_{\rm II}^{(0)}+n_{\rm I}^{(\hat{\sigma})},\quad \delta f_2^{(\sigma)}=\delta n_{\rm I}^{(\hat{\sigma})}=-\delta n_{\rm I}^{(\sigma)}, \tag{38}$$

which are actually the same as those in the two- and one-particle d-states.

From this we conclude at once that the general Eqs. (28), which are independent of the magnetic field variation, are unchanged:

$$\delta n_{\rm II}^{\sigma}(1-K_p)-2K_p\delta n_{\rm I}^{(\sigma)}=\delta f_1^{(\sigma)}(1-K_p)+\delta f_2^{(\sigma)}(1+K_p)=0. \tag{39}$$

The equation for the susceptibility, expressed in terms of the variations in the end factors, contains twice the number of terms on the left and right. Thus, the final equation contains an extra factor of 1/2 multiplying the variation δf_e of the one-particle e-states:

$$\delta n_{\rm II}^{(\sigma)}=\delta f_2^{(\sigma)}+\delta f_1^{(\sigma)}=[K_p+L_p]\sum_{k=1,2}g_k^2\delta f_k^{(\sigma)}+g_p^2\frac{f_p}{2f_e}L_p\delta f_e-g_p^2f_pR_p\sigma\delta H. \tag{40}$$

The difference shows up only when calculating the number of identical transitions, which is twice as great. The sum of the squares of the genealogical coefficients is also doubled: $g_p^2=3$.

By analogy with Eq. (14) we find an equation for the variation in f_e with a coefficient that is twice as large for the variations $\delta f_{1,2}$:

$$\delta f_e^{(\sigma)}=[K_e+L_e]\delta f_e^{(\sigma)}+2\frac{f_e}{g_p^2f_p}L_d\sum_{k=1,2}g_k^2\delta f_k^{(\sigma)}-f_eR_e\sigma\delta H. \tag{41}$$

Here all the coefficients are determined for $H=0$ using the same formulas, but with different equations for K_p and K_e. We have

$$K_p=2\frac{n_p-1}{4-n_p},\quad K_e=\frac{n_e}{(4-3n_e)},$$
$$n_p=2\frac{(2K_p+1)}{(2+K_p)},\quad n_e=\frac{4K_e}{(1+3K_e)}. \tag{42}$$

The condition for a ferromagnetic instability has the same form (35), but with transposed indices, $p\leftrightarrow e$:

$$(1-K_e)[K_p(1-K_p)-L_p(1/3+K_p)]-L_eK_p(1-K_p)=0\quad\text{for } n_e<1,\quad 1<n_p<2. \tag{43}$$

The common property of Eqs. (36) and (43), which distinguishes them from Eq. (16), is the factor K_e and, therefore, K_p, which reflects the possible onset of ferromagnetism for a small number of excitations when $n_e-1\ll 1$ or $n_p-1\ll 1$.

Concentrations $1<n_p<2$, $1<n_e<2$. We now obtain the equations appropriate to the range $1<n_{p,e}<2$ by generalizing the results of the preceding two sections. Equations (28), (38), and (39) remain unchanged.

We find two other equations for the two-particle occupation numbers $n^{\sigma}_{p,e}$ by analogy with Eqs. (31) and (40):

$$\delta n^{(\sigma)}_{\mathrm{II},e}=\delta f^{(\sigma)}_{2,e}+\delta f^{(\sigma)}_{1,e}=[K_e+L_e]\sum_{\alpha=1,2} g^2_{\alpha}\delta f^{(\sigma)}_{\alpha,e} + 2g^2_e\frac{f_e}{g^2_p f_p}L_e\sum_{k=1,2} g^2_k\delta f^{\sigma}_{k,p}-g^2_e f_e R_e\sigma\delta H, \quad (44)$$

and

$$\delta n^{(\sigma)}_{\mathrm{II},p}=\delta f^{(\sigma)}_{2,p}+\delta f^{(\sigma)}_{1,p}=[K_p+L_p]\sum_{k=1,2} g^2_k\delta f^{(\sigma)}_{k,p} + g^2_p\frac{f_p}{2g^2_e f_e}L_p\sum_{\alpha=1,2}\delta g^2_{\alpha=1,2} f^{\sigma}_{\alpha,d}-g^2_p f_p R_p\sigma\delta H. \quad (45)$$

Here $g^2_p=3$, $g^2_e=3/2$, $K_e=4(n_e-1)/(2+n_e)$, and $K_p=2(n_p-1)/(4-n_p)$.

The equation for the ferromagnetic instability boundary is symmetric under the interchange $p\leftrightarrow e$:

$$K_pK_e(1-K_p)(1-K_e)=L_pK_e(1-K_e)(1/3+K_p) + L_eK_p(1-K_p)(1/3+K_e). \quad (46)$$

Here

$$K_p=2\frac{n_p-1}{4-n_p},\quad K_e=4\frac{(n_e-1)}{2+n_e},$$
$$n_p=2\frac{(2K_p+1)}{(2+K_p)},\quad n_e=2\frac{K_e+2}{4-K_e}. \quad (47)$$

For positive $L_{p,e}\approx 1$ and a small number of perturbations, $n_{p,e}-1\ll 1$, the left-hand side of Eq. (46) is small compared to the right, which reflects the possible onset of ferromagnetism.

Concentrations $n_e<1$, $2<n_p<3$. We now examine the most complicated case, in which the *p*-electrons resonate between two- and three-particle states. We assume that $2<n_p<3$, while the occupation numbers n_e are less than unity.

The lowest three-particle state has $S=3/2$ and a fourfold degeneracy with respect to the projection of the spin:

$$\hat a^+_{\sigma}\hat b^+_{\sigma}\hat c^+_{\sigma}|0\rangle,\quad S_z=3\sigma/2,$$
$$\frac{1}{\sqrt3}(\hat a^{\pm}_{\bar\sigma}\hat b^+_{\sigma}\hat c^+_{\sigma}|0\rangle+\hat a^+_{\sigma}\hat b^{\pm}_{\bar\sigma}\hat c^+_{\sigma}|0\rangle+\hat a^+_{\sigma}\hat b^+_{\sigma}\hat c^{\pm}_{\bar\sigma}|0\rangle),\quad S_z=\frac{\sigma}{2}. \quad (48)$$

The three lowest triplet states with spin 1 are constructed from three different pairwise products of creation operators [see the definition (21)].

The expansion in *X*-operators of the transition between the lowest energy two- and three-particle states is determined by the three genealogical coefficients ($\hat a_{\mathbf{r}\sigma}=\hat p_x$):

$$\hat a_{\mathbf{r}\sigma}=\hat X^{(0,\sigma,\sigma|3\sigma/2)}_{\mathbf{r}}+\sqrt{\frac{2}{3}}\hat X^{(A(yz,xz)|\sigma/2)}_{\mathbf{r}}+\frac{1}{\sqrt3}\hat X^{(0,\bar\sigma,\sigma|\bar\sigma/2)}_{\mathbf{r}}. \quad (49)$$

We obtain the expansion of the two other annihilation operators $\hat b=\hat p_y$ and $\hat c=\hat p_z$ from Eq. (49) by cyclic permutation.

In the absence of a field, all the average occupation numbers and end factors can be expressed in terms of n_p, the average number of electrons per cell. Given the order of the degeneracy, we have

$$3n_{\mathrm{II}}+4n_{\mathrm{III}}=1,\quad 18n_{\mathrm{II}}+12n_{\mathrm{III}}=n_p,\quad f_p=\frac{5n_p-6}{36}. \quad (50)$$

Summing over the spin index, we obtain the equation of state

$$n_p=2+4f_pK_p,$$
$$K_p=\frac{1}{3}\left\{n_F(\epsilon_p)+\sum_{\mathbf{p},\lambda=\pm}a^{\lambda}_{\mathbf{p}}n_F(\xi^{(\lambda)}_{\mathbf{p}})\right\},$$
$$\xi^{(\pm)}_{\mathbf{p}}=\pm\sqrt{\left(\frac{r}{2}\right)^2+2\,f_pf_et^2_{\mathbf{p}}}-\mu. \quad (51)$$

To find the equation of state for $H\neq 0$, we write the occupation number of the three-particle states in terms of the one particle Green function at the same points.

The equations for the variations in the three-particle occupation numbers $\delta n^{(3\sigma/2)}_{\mathrm{III}}$ and $\delta n^{(\sigma/2)}_{\mathrm{III}}=-\delta n^{(-\sigma/2)}_{\mathrm{III}}$ can be obtained from the general equation for the average *T*-product of the annihilation operator (49) and a linear combination of the three conjugate operators with arbitrary coefficients γ:

$$g_1\gamma_1 n^{(3\sigma/2)}_{\mathrm{III}}+g_2\gamma_2 n^{(\sigma/2)}_{\mathrm{III}}+g_3\gamma_3 n^{(-\sigma/2)}_{\mathrm{III}} = T\sum_{1\leqslant k,n\leqslant 3}\sum_{\omega,\mathbf{p},k} g_kG^{k,n}_{\omega}(\mathbf{p})\gamma_nf_n\exp(i\omega\delta). \quad (52)$$

The matrix elements of the one-particle Green function can be expressed in terms of the inverse matrix, where now each row and column, reflecting the *p*-states, has three times as many components. The corresponding matrix elements are determined by the genealogical coefficients g_k in the expansion (49). Thus, for the selected operator $\hat a=\hat p_{x\sigma}$, the three first rows of the inverse Green function can be written

$$\hat G^{-1}_{\omega}(\mathbf{p})=\begin{pmatrix}\left(\frac{1}{2}\sigma\rightarrow\frac{3}{2}\sigma\right);\\ \left(-\frac{1}{2}\sigma\rightarrow\frac{1}{2}\sigma\right);\\ \left(-\frac{3}{2}\sigma\rightarrow-\frac{1}{2}\sigma\right);\end{pmatrix}$$

$$\begin{array}{c} \\ k=1 \\ k=2 \\ k=3 \\ k=4,5.... \end{array} \begin{array}{c} \begin{array}{cccc} s=1 & s=2 & s=3 & s=4,5,... \end{array} \\ \begin{pmatrix} i\omega-\epsilon_p & 0 & 0 & -f_1 g_1 \rho_s^p \\ 0 & i\omega-\epsilon_p & 0 & -f_2 g_2 \rho_s^p \\ 0 & 0 & i\omega-\epsilon_p & -f_3 g_3 \rho_s^p \\ -\nu_p^k g_1 & -\nu_p^k g_2 & -\nu_p^k g_3 & \tau_{k,s} \end{pmatrix} \end{array}. \tag{53}$$

Here $g_1=1$, $g_2=\sqrt{2/3}$, and $g_3=\sqrt{1/3}$, while the end factors are determined by the sum of the occupation numbers for the two and three-particle states:

$$f_1^\sigma = n_{\mathrm{III}}^{(3\sigma/2)} + n_{\mathrm{II}}^\sigma, \quad f_2^\sigma = n_{\mathrm{III}}^{(\sigma/2)} + n_{\mathrm{II}}^0,$$

$$f_1^\sigma = n_{\mathrm{III}}^{(-\sigma/2)} + n_{\mathrm{II}}^{-\sigma}.$$

Using the explicit form of the elements of the inverse matrix, it is possible to calculate the sums on the right-hand side of Eq. (53). Since the energies of the transitions accompanying the creation of the same p_x-state are the same, i.e., $\epsilon_1=\epsilon_2=\epsilon_3=\epsilon_p$, we have

$$\sum_{k=1}^{3} g_k G_\omega^{k,s}(\mathbf{p}) = g_s (i\omega-\epsilon_p)^2 \frac{\det \tau_{n,m}}{\det G^{-1}}. \tag{54}$$

Summing over momentum and frequency on both sides of this equation, we obtain a result proportional to the genealogical coefficients g_s:

$$T\sum_{\omega,\mathbf{p}} \sum_{k=1}^{3} g_k G_\omega^{k,s}(\mathbf{p}) = g_s K_p(H).$$

As a result we have an equation for arbitrary γ_k:

$$g_1\gamma_1 n_{\mathrm{III}}^{(3\sigma/2)} + g_2\gamma_2 n_{\mathrm{III}}^{(\sigma/2)} + g_3\gamma_3 n_{\mathrm{III}}^{(-\sigma/2)} = K_p(H) \sum_{s=1,2,3} g_s\gamma_s f_s. \tag{55}$$

If we assume that the vector γ is orthogonal to the vector g, i.e.,

$$\sum_{1\leqslant k\leqslant 3} g_k\gamma_k = 0,$$

then it is possible to obtain two equations that do not depend explicitly on the magnetic field variation.

We find the first equation for the condition $g_1\gamma_1 = g_3\gamma_3$, $g_2\gamma_2 = -2g_1\gamma_1$:

$$\delta n_{\mathrm{III}}^{(3/2)} = 3\,\delta n_{\mathrm{III}}^{(1/2)}. \tag{56}$$

If, on the other hand, we set $\gamma_2=0$ and $g_3\gamma_3 = -g_1\gamma_1$, then we have the second equation:

$$(1-K_p)(\delta n_{\mathrm{III}}^{(3/2)} + \delta n_{\mathrm{III}}^{(1/2)}) - 2K_p\,\delta n_{\mathrm{II}} = 0. \tag{57}$$

The variations in the end factors can be expressed in terms of the variations in the occupation number:

$$\delta f_1 = \delta n_{\mathrm{III}}^{(3/2)} + \delta n_{\mathrm{II}}, \quad \delta f_2 = \delta n_{\mathrm{III}}^{(1/2)},$$

$$\delta f_3 = \delta n_{\mathrm{III}}^{(-1/2)} - \delta n_{\mathrm{II}}. \tag{58}$$

Using the additional condition $\delta n_{\mathrm{III}}^{(-1/2)} = -\delta n_{\mathrm{III}}^{(1/2)}$, we find the inverse relations:

$$\delta n_{\mathrm{III}}^{(3/2)} = \delta f_1 + \delta f_2 + \delta f_3, \quad \delta n_{\mathrm{III}}^{(1/2)} = \delta f_2;$$

$$\delta n_{\mathrm{II}} = -\delta f_2 - \delta f_3. \tag{59}$$

The variation of the virtual Green function $\delta K_p(H)$ contains three types of terms:

$$g_1^2\delta n_{\mathrm{III}}^{(3\sigma/2)} + g_2^2\delta n_{\mathrm{III}}^{(\sigma/2)} + g_3^2\delta n_{\mathrm{III}}^{(-\sigma/2)} = [K_p+L_p] \sum_{k=1,2,3} g_k^2 \delta f_k^{(\sigma)} + g^2 \frac{f_e}{f_p} L_e\,\delta f_e - g^2 f_p R_p \sigma\,\delta H, \tag{60}$$

where the coefficients $L_{p,e}$ are determined by the general formula (8)

$$R_p = \frac{1}{3}\left[n_F'(\epsilon_p) + 2 \sum_{\mathbf{p},\lambda=\pm} a_{\mathbf{p}}^{(\lambda)} n_F'(\xi_{\mathbf{p}}^{(\lambda)}) \right],$$

$$a_{\mathbf{p}}^{(\pm)} = \frac{1}{2}\left[1 \pm \frac{r}{\sqrt{r^2 + 4g^2 f_e f_p \tau_{\mathbf{p}}^2}} \right]. \tag{61}$$

Thus, we obtain the same equations as Eqs. (15)–(17), but with different definitions of the parameters:

$$g^2=2, \quad g_1^2=1, \quad g_2^2=\frac{2}{3}, \quad g_3^2=\frac{1}{3}, \quad K_p = 9\,\frac{n_p-2}{5n_p-6}. \tag{62}$$

The equation for the variation in n_e is analogous in form to Eqs. (22) and (23):

$$\delta n_d^{(\sigma)} = 2\,\delta f_e^{(\sigma)} = L_e \frac{f_e}{g^2 f_p} \sum_{k=1,2,3} g_k^2 \delta f_k^{(\sigma)} + [K_e+L_e]\delta f_e - f_d R_d \sigma\,\delta H. \tag{63}$$

All coefficients are calculated for zero magnetic field:

$$K_e = \frac{n_e}{4-3n_e}, \quad R_e = \sum_{\mathbf{p},\lambda=\pm} a_{\mathbf{p}}^{(\lambda)} n_F'(\xi_{\mathbf{p}}^{(\lambda)}). \tag{64}$$

Thus, we have a system of four equations from which we can find the condition for the onset of ferromagnetism:

$$\det\begin{pmatrix} 1-g_1^2(K_p+L_p) & 4/3-g_2^2(K_p+L_p) & 1-g_3^2(K_p+L_p) & -g^2 f_p L_p/f_e \\ -1 & +2 & -1 & 0 \\ 1-K_p & 2 & 1+K_p & 0 \\ -g_1^2 f_e L_e/g^2 f_p & -g_2^2 f_e L_p/g^2 f_p & -g_3^2 f_e/g^2 f_p & 1-K_e-L_e \end{pmatrix}=0. \tag{65}$$

Calculating the determinant yields the following condition:

$$(1-K_e)[K_p(1-K_p)-L_p(2/3+K_p)]-L_e K_p(1-K_p)=0$$
$$\text{for } n_e<1, \quad 1<n_p<2. \tag{66}$$

Here

$$K_p=9\frac{(n_p-2)}{5n_p-6}, \quad K_e=\frac{n_e}{4-3n_e},$$
$$n_p=6\frac{3-K_p}{9-5K_p}, \quad n_e=4\frac{K_e}{1+3K_e}. \tag{64′}$$

A comparison of Eqs. (66) and (43) shows that upon going from a resonance between one- and two-particle states to a resonance between two- and three-particle states, there is simply an increase in the intensity of the effective scattering amplitude, since in the latter case the dimensionless amplitude L_p is multiplied by 2/3, twice the value of 1/3 obtained in the former case.

Concentrations $1<n_e<2$, $2<n_p<3$. The matching equations (41), when written for the variations in the three particle *p*-states, acquire an additional term associated with the possible variation in the one- and two-particle *e*-states:

$$g_1^2\delta n_{\rm III}^{(3\sigma/2)}+g_2^2\delta n_{\rm III}^{(\sigma/2)}+g_3^2\delta n_{\rm III}^{(-\sigma/2)}$$
$$=[K_p+L_p]\sum_{k=1,2,3} g_k^2\delta f_k^{(\sigma)}+g_p^2\frac{f_p}{g_e^2 f_e}L_d$$
$$\times\sum_{\alpha=1,2}\delta g_{\alpha,e}^2 f_{\alpha,e}^{\sigma}-g_p^2 f_p R_p\sigma\delta H. \tag{67}$$

Equations (37) and (38) for the coupling between the variations in the three and two-particle *p*-states are unchanged.

The equation of the type (31) for the variation in the two-particle *d*-states acquires three new terms on the right:

$$\delta n_{\rm II}^{(\sigma)}=\delta f_{2,e}^{(\sigma)}+\delta f_{1,e}^{(\sigma)}=[K_e+L_e]\sum_{\alpha=1,2} g_\alpha^2\delta f_{\alpha,e}^{(\sigma)}$$
$$+g_e^2\frac{f_e}{g_p^2 f_p}L_e\sum_{k=1,2,3}\delta f_{k,p}^{\sigma}-g_e^2 f_e R_e\sigma\delta H. \tag{68}$$

Equation (28), which relates the variations in the one-and two-particle *p*-states, is unchanged.

We obtain the condition a condition that is fully analogous to Eq. (46), but with a doubling of the coefficient of the dimensionless amplitude L_p:

$$K_p K_e(1-K_p)(1-K_e)$$
$$=L_p K_e(1-K_e)(2/3+K_p)+L_e K_p(1-K_p)(1/3+K_d), \tag{69}$$

with

$$K_p=9\frac{(n_p-2)}{(5n_p-6)}, \quad K_e=4\frac{(1-n_e)}{(2+n_e)},$$
$$n_p=6\frac{3-K_p}{9-5K_p}, \quad n_e=2\frac{K_e+2}{4-K_e}. \tag{70}$$

Concentrations $2<n_e<4$, $0<n_p<3$, and $0<n_e<4$, $3<n_p<6$. For the transition into the region $2<n_e<4$, $0<n_p<2$ we can use the equation from the preceding sections with a particle-hole symmetry transformation for the e_g-electrons: $n_e\to(4-n_e)$, $K_e\to(1-K_e)$, $L_e\to-L_e$.

The first two transformations correspond to a transformation from particles to holes for the fourfold degenerate e_g-shell. The last transformation corresponds to a change in the sign of the *e*-*e* scattering amplitude upon going from particles to holes, which leads to a major change in the magnetic phase diagram.

The remaining part of the phase diagram for $0<n_e<4$, $3<n_p<6$ can be obtained from the formulas of the preceding sections using the general particle-hole symmetry transformations $n_d\to4-n_d$ and $n_p\to6-n_p$.

3. POWER-LAW DENSITY OF STATES MODEL

Let us calculate the integrals (17) and (18) for a model with a density of states $\rho(\epsilon)=(2/3)\sqrt{\epsilon}\,\theta(\epsilon)\,\theta(1-\epsilon)$. In the case of a filled lower subband, for which the chemical potential changes in the negative region from $-|r/2|\sqrt{1+s}$ to $-|r/2|$, where $s=12g_e^2 g_p^2 f_e f_p t^2/r^2$, for $T=0$ we have

$$K_p=\frac{1}{3}\left\{1-z^{3/2}-\frac{3\,\mathrm{sign}(r)}{2}W(s,z)\right\},$$
$$K_e=\frac{1}{2}\left\{1-z^{3/2}+\frac{3\,\mathrm{sign}(r)}{2}W(s,z)\right\}. \tag{71}$$

Henceforth we make use of the function

$$W(s,z)=\frac{\sqrt{1+s}-\sqrt{z(1+sz)}}{s}$$
$$-\frac{1}{s\sqrt{s}}\ln\left(\frac{\sqrt{s}+\sqrt{1+s}}{\sqrt{zs}+\sqrt{1+sz}}\right). \tag{72}$$

Instead of a negative chemical potential μ here we have introduced the notation $-\mu=\sqrt{1+sz}$.

For this power-law dependence of the density states, the functions $L_{p,e}$ are given in terms of $K_{p,e}$ by

$$L_p=\frac{1}{2}-\frac{\mathrm{sign}(r)}{2\sqrt{1+s}}-\frac{3}{2}K_p, \quad L_e=\frac{3}{4}+\frac{3\,\mathrm{sign}(r)}{4\sqrt{1+s}}-\frac{3}{2}K_e. \tag{73}$$

TABLE I.

Ranges	f_p	K_p	g_p^2	$S_{[n_p]}$	$R_{[n_p]+1}$	Γ_p	SS
$0<n_p<1$	$(6-5n_p)/6$	$n_p/(6-5n_p)$	1	0	6	0	t_{2g}
$1<n_p<2$	$(4-n_p)/18$	$2(n_p-1)/(4-n_p)$	3	1/2	9	1/3	t_{2g}^2
$2<n_p<3$	$(5n_p-6)/36$	$9(n_p-2)/(5n_p-6)$	2	1	4	2/3	t_{2g}^3
$3<n_p<4$	$(24-5n_p)/36$	$4(n_p-3)/(24-5n_p)$	2	3/2	9	$-5/3$	t_{2g}^4
$4<n_p<5$	$(n_p-2)/18$	$3(n_p-4)/(n_p-2)$	3	1	6	$-4/3$	t_{2g}^5
$5<n_p<6$	$(5n_p-24)/6$	$6(n_p-5)/(5n_p-24)$	1	1/2	1	-1	t_{2g}^6
Ranges	f_e	K_e	g_e^2	$S_{[n_e]}$	$R_{[n_e]+1}$	Γ_e	SS
$0<n_e<1$	$(4-3n_e)/4$	$n_e/(4-3n_e)$	1	1/2	4	0	e_g
$1<n_e<2$	$(2+n_e)/12$	$4(n_e-1)/(2+n_e)$	3/2	1/2	3	1/3	e_g^2
$2<n_e<3$	$(6-n_e)/12$	$3(n_e-2)/(6-n_e)$	3/2	1	4	$-4/3$	e_g^3
$3<n_e<4$	$(3n_e-8)/4$	$4(n_e-3)/(3n_e-8)$	1	1/2	4	-1	e_g^4

Here $s>0$ and the parameter z varies from zero to one. The amplitudes $L_{p,e}$ remain positive for all s and z.

In a filled upper subband, for which the chemical potential varies over the positive range from $|r/2|$ to $|r/2|\sqrt{1+s}$, we have

$$K_p=\frac{1}{3}\left\{2+z^{3/2}-\frac{3\,\mathrm{sign}(r)}{2}W(s,z)\right\},$$

$$K_e=\frac{1}{2}\left\{1+z^{3/2}+\frac{3\,\mathrm{sign}(r)}{2}W(s,z)\right\}. \tag{74}$$

and

$$L_p=1-\frac{\mathrm{sign}(r)}{2\sqrt{1+s}}-\frac{3}{2}K_p,$$

$$L_e=\frac{3}{4}+\frac{3\,\mathrm{sign}(r)}{4\sqrt{1+s}}-\frac{3}{2}K_e. \tag{75}$$

Here also, the parameter z varies from zero to one, but the dimensionless amplitude $L_{p,e}$ remains negative.

Hence, we conclude that within the range of occupation numbers $0<n_p<2$ and $0<n_e<2$, ferromagnetic regions can exist only for negative values of the chemical potential (see Fig. 2).

4. GENERAL RESULTS AND CONCLUSIONS

All of our results can be summarized by a single equation with constant coefficients, each of which has a certain meaning within a square domain of the variables $n_{p,d}$:

$$K_p(1-K_p)K_e(1-K_e)$$
$$=L_e(\Gamma_e+K_e)K_p(1-K_p)+L_pK_e(1-K_e)(\Gamma_p+K_p). \tag{76}$$

All coefficients are defined in Table I.

The results in the table show that changing the sign of the e-e scattering amplitude within the range $2<n_e<4$ for a fixed sign of the p-p scattering amplitude in the range $0<n_p<3$ causes them largely to cancel. As a result, the right-hand side of Eq. (76) becomes smaller, while the left remains unchanged, since it is invariant under particle-hole transformations for p- and e-electrons individually.

For this reason, hybridization of states from the lower Hubbard subbands $n_p<3$ with states from the upper half of the e-subband $2<n_e$ does not lead to ferromagnetism (see Fig. 2).

The figure shows that for the chosen initial density of states, ferromagnetic instability does not occur at low $n_{p,e}<1$ either. With a negative chemical potential, the entire system is spread out over a large number of vacant (nonmagnetic) states and a small number of one-particle states with spin 1/2. Thus, in this region a gaseous paramagnetic state does exist, analogous to that in the classical Hubbard model.[2]

In the other regions where, for example, $n_e>1$, the system is a mixture of magnetic states. The existence of a finite negative scattering amplitude for excitations with the same sign of the projection of the spin (or a positive amplitude for scattering of excitations with opposite spins) leads to ferromagnetic instability even when the number of excited states is small. A similar result has been obtained elsewhere.[6]

This sort of situation arises during hybridization and simultaneous filling of the lower Hubbard p- and e-subbands.

If, on the other hand, the upper subband for the e-states and the lower subband for the p-states are filled simultaneously, then the physical situation becomes more complicated.

Calculations show that when the lower half of a hybridized subband (negative chemical potential) is filled, we see a negative $e-e$ scattering amplitude and a positive $p-p$ scattering amplitude for excitations with opposite spins. When the upper half of the hybridized subband is filled (positive chemical potential), the situation is the opposite: we see a positive $e-e$ scattering amplitude and a negative $p-p$ scattering amplitude for excitations with opposite spins.

The mutual influence of these interactions is what determines the possible onset of ferromagnetism. The calculations done in this paper indicate that the amplitudes largely cancel for arbitrary concentrations n_p and n_e of the states (see the right half of Fig. 2).

For comparison with experiments on the ferromagnetic fcc compounds FeAl and MnSi, we note that in this case, because of electrical neutrality, the total number of electrons in unfilled shells in nine: $n_d+n_p+n_s=9$. Here n_d is the total number of *d*-electrons per cell, which differs from our n_e by 6, the total number of electrons in the filled t_{2g}-cell.

In fact, n_s is a fit parameter. According to band calculations,[5] it is 0.84 and 0.82, respectively, for iron and manganese.

Substituting these numbers into the condition for electrical neutrality, we find

$$n_e+n_p=2.16 \quad \text{for FeAl,} \quad n_e+n_p=2.18 \quad \text{for MnSi.} \tag{77}$$

In the figure these lines essentially overlap, with the electrical neutrality line for FeAl lying slightly below that for MnSi. Both of these lines cross the two shaded regions of ferromagnetic ordering obtained in our model.

Suppose that the average charge of the anions is low: for Al $n_p\approx 1$ and for Si $n_p\approx 2$. Then $n_e\approx 1$ for FeAl, while $n_e\ll 1$ for MnSi. Therefore, we obtain qualitative agreement with experiment: the magnetic moment in MnSi is small and the saturation moment in the ferromagnetic material FeAl is large (≈ 1).[7]

Note that the proposed mechanism for the onset of the ferromagnetic instability allows for the possibility of intersection of the phase transformation boundary and the electrical neutrality line. This situation corresponds to the phenomenological theory of Kittel,[8] according to which the exchange interaction constant changes sign at some critical volume of the unit cell.[9]

There is yet another possibility, in which the electrical neutrality line intersects the line corresponding to integer values of $n_k=[n_k]$, as happens in our case. It can be shown, however, that the region $(n_k-[n_k])\leqslant|t|/U$, within which a transition to localized moments takes place, abuts this line. The assumption in this paper of infinite Hubbard energy U makes it impossible to examine this region, and physical phenomena within this narrow range, $(n_k-[n_k])\ll|t|/U$, warrant special treatment.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 98-02-17388).

*)E-mail: zaitsev@mbslab.kiae.ru; zam@niitp.mainet.msk.su

[1] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).
[2] J. Hubbard, Proc. Roy. Soc. A **276**, 238 (1963).
[3] R. O. Zaĭtsev, Zh. Éksp. Teor. Fiz. **70**, 1100 (1976) [Sov. Phys. JETP **43**, 574 (1976)].
[4] J. Hubbard and K. P. Jain, J. Phys. C **2**, 1650 (1968).
[5] W. Harrison, *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond*, Dover, New York (1989).
[6] T. Momoi and K. Kubo, Phys. Rev. B **58**, R567 (1998).
[7] I. S. Grigor'ev and E. Z. Meĭlikhov (Eds.), *Physical Quantities* [in Russian], Énergoatomizdat, Moscow (1991), Ch. 27.
[8] C. Kittel, Phys. Rev. **120**, 335 (1960).
[9] C. P. Bean and D. S. Rodbell, Phys. Rev. **126**, 104 (1962).

Translated by D. H. McNeill

Superdiffusion and stable laws

V. M. Zolotarev

V. A. Steklov Institute of Mathematics, Russian Academy of Sciences, 117966, Moscow

V. V. Uchaĭkin*) and V. V. Saenko

Ulyanovsk State University, 432700, Ulyanovsk

(Submitted 20 March 1998)
Zh. Éksp. Teor. Fiz. **115**, 1411–1425 (April 1999)

The superdiffusion equation with a fractional Laplacian $\Delta^{\alpha/2}$ in N-dimensional space describes the asymptotic ($t\to\infty$) behavior of a generalized Poisson process with the range (discontinuity) distribution density $\sim|x|^{-\alpha-1}$. The solutions of this equation belong to a class of spherically symmetric stable distributions. The main properties of these solutions are given together with their representations in the form of integrals and series and the results of numerical calculations. It is shown that allowance for the finite velocity of free particle motion for $\alpha>1$ merely amounts to a reduction in the diffusion coefficient with the form of the distribution remaining stable. For $\alpha<1$ the situation changes radically: the expansion velocity of the diffusion packet exceeds the velocity of free particle motion and the superdiffusion equation becomes physically meaningless. © *1999 American Institute of Physics.*
[S1063-7761(99)01804-1]

1. INTRODUCTION

In view of the increasing interest being shown in anomalous diffusion processes,[1,2] there has recently been intensive discussion of the possibility of generalizing the diffusion equation

$$\frac{\partial\rho(\mathbf{r},t)}{\partial t}=D\Delta\rho(\mathbf{r},t),\quad \rho(\mathbf{r},t)=\delta(\mathbf{r}) \tag{1}$$

by replacing the ordinary differentiation operators with corresponding fractional-order operators. From the physical point of view, the substitution $\partial/\partial t\to(\partial/\partial t)^{\beta}$, $\beta<1$ is caused by the influences of traps distributed in a medium in which the random time τ of particle residence is characterized by the probability density

$$q(t)\propto t^{-\beta-1},\quad t\to\infty$$

with an infinite mean, and the introduction of the fractional Laplacian $\Delta^{\alpha/2}$ is associated with an anomalously broad distribution of particle ranges

$$p(r)\propto r^{-\alpha-1},\quad r\to\infty.$$

This first regime is called subdiffusion (the diffusion packet spreads more slowly with time than in cases of ordinary diffusion) and the second is called superdiffusion (where the packet spreads more rapidly).[3–6] We shall confine our analysis to the second anomalous diffusion regime, having noted that an equation with a fractional-order Laplacian ($\alpha=1/3$) was first proposed by Monin[7] (see also Ref. 8) in a description of diffusion in a turbulent medium.

Compte[9] proposed a generalization of Eq. (1) to the case of superdiffusion in the form

$$\frac{\partial\rho}{\partial t}=D\nabla^{\alpha}\rho(\mathbf{r},t)$$

together with

$$\hat{\mathsf{F}}\nabla^{\alpha}\rho=k^{\alpha}\hat{\rho}(\mathbf{k},t), \tag{2}$$

where

$$\hat{\mathsf{F}}\,f\equiv\hat{f}\mathbf{k}=\int e^{i\mathbf{k}\cdot\mathbf{r}}f(\mathbf{r})\,d\mathbf{k}$$

indicates the Fourier transform of the function $f(\mathbf{r})$ and the exponent α is extended to the entire interval (0,2). However, we note that for $\alpha\to2$ formula (2) does not yield the result

$$\hat{\mathsf{F}}\nabla^{2}\rho=-k^{2}\hat{\rho},$$

valid for normal diffusion and for $\alpha=1$ the operator ∇^{1}, differs from the vector operator ∇ which makes it difficult to interpret the fractional differentiation operator as the fractional exponent of the differential operator.

These problems are easily surmounted by determining the Riesz fractional derivative by a standard method, in terms of the Laplacian, as in Ref. 10, which incidentally is also quoted by Compte.[9] The superdiffusion equation then has the form

$$\frac{\partial\rho}{\partial t}=-D(-\Delta)^{\alpha/2}\rho(\mathbf{r},t), \tag{3}$$

where the operator $(-\Delta)^{\alpha/2}$ is given by

$$(-\Delta)^{\alpha/2}\rho=\hat{\mathsf{F}}^{-1}|\mathbf{k}|^{\alpha}\hat{\mathsf{F}}\rho. \tag{4}$$

The definition (4) remains valid for arbitrary dimensions N of the space, whose vectors will be denoted by x and k and for which the scalar product $k_1x_1+\ldots+k_Nx_N$ will be denoted by kx, so that the Fourier transform of the function $f(x)$, $x\in\mathsf{R}^N$ is written as

$$\hat{\mathsf{F}}f \equiv \hat{f}(k) = \int e^{ikx} f(x)\, dx.$$

The fractional Laplacian is expressed explicitly in the form of the hypersingular integral

$$(-\Delta)^{\alpha/2} f = \frac{1}{d_{N,n}(\alpha)} \int_{\mathsf{R}^N} \frac{(\Delta_y^n f)(x)}{|y|^{N+\alpha}} dy,$$

where

$$(\Delta_y^n f)(x) = \sum_{m=0}^{n} (-1)^m \binom{n}{m} f(x - my)$$

are noncentrosymmetric differences, n is any integer greater than α, and the normalization factor $d_{N,n}(\alpha)$ is given by

$$d_{N,n}(\alpha) = \int_{\mathsf{R}_N} (1 - \exp ix_1)^n |x|^{-N-\alpha}\, dx.$$

The Riesz theory of differentiation is expounded in detail in Ref. 10, which is a unique encyclopedia of fractional analysis.

In the present paper we discuss the physical meaning of Eq. (3), its solutions, and range of validity. Bearing in mind the independent importance of one-dimensional and two-dimensional diffusion problems, we shall consider N-dimensional diffusion.

2. LÉVY PROCESS

Superdiffusion can easily be described formally in terms of a generalization of the Wiener process. For this purpose it is sufficient to write the Kolmogorov–Chapman equation for a steady-state Markovian process with independent increments

$$\rho(x,t) = \int \rho(x',t')\rho(x-x',t-t')\, dx', \tag{5}$$

$$x, x' \in \mathsf{R}^N, \quad t, t' \in \mathsf{R}^1_+, \quad 0 < t' < t, \quad \rho(x,0) = \delta(x),$$

and analyze the class of its self-similar solutions

$$\rho^{(\alpha)}(x,t) = (Dt)^{-N/\alpha} g^{(\alpha)}(x(Dt)^{-1/\alpha}), \quad D > 0, \quad \alpha > 0. \tag{6}$$

A Fourier transformation converts Eqs. (5) and (6) to the form

$$\hat{\rho}^{(\alpha)}(k,t) = \hat{\rho}^{(\alpha)}(k,t')\hat{\rho}^{(\alpha)}(k,t-t'),$$

$$\hat{\rho}^{(\alpha)}(k,t) = \hat{g}^{(\alpha)}(k(Dt)^{1/\alpha}), \quad k \in \mathsf{R}^N.$$

Introducing the notation $(Dt)^{1/\alpha} = \lambda$, $(Dt')^{1/\alpha} = \lambda_1$, and $(D(t-t'))^{1/\alpha} = \lambda_2$, we can see that $\hat{g}^{(\alpha)}(k)$ satisfies

$$\ln \hat{g}^{(\alpha)}(\lambda k) = \ln \hat{g}^{(\alpha)}(\lambda_1 k) + \ln \hat{g}^{(\alpha)}(\lambda_2 k) \tag{7}$$

for

$$\lambda^\alpha = \lambda_1^\alpha + \lambda_2^\alpha. \tag{8}$$

For spherically symmetric distributions, $\hat{g}^{(\alpha)}(k)$ only depends on $|k|$ and then the solution of Eq. (7) is (see Ref. 11)

$$\hat{g}^{(\alpha)}(k) = e^{-|k|^\alpha}, \quad 0 < \alpha \leqslant 2, \tag{9}$$

and in consequence

$$\hat{\rho}^{(\alpha)}(k,t) = e^{-D|k|^\alpha t} \tag{10}$$

The constraint $\alpha \leqslant 2$ is imposed because for $\alpha > 2$ the function (9) loses the properties of a characteristic function, i.e., its inverse Fourier transform ceases to be a probability density (it becomes indefinite).

The Fourier transform (10) satisfies the differential equation

$$\frac{\partial \hat{\rho}^{(\alpha)}}{\partial t} = -D|k|^\alpha \hat{\rho}^{(\alpha)},$$

which, in accordance with an N-dimensional analog of formula (4), as a result of an inverse transformation, yields an equation for the probability density

$$\frac{\partial \rho^{(\alpha)}}{\partial t} = -D(-\Delta)^{\alpha/2} \rho^{(\alpha)}(x,t), \quad \rho(x,t) = \delta(x), \tag{11}$$

which generalizes Eq. (3) to the N-dimensional case. In the one-dimensional case, the operator $(-\Delta)^{\alpha/2}$ becomes a Marchaux derivative.[10] Following Montroll and West,[12] we shall call this process a Lévy process. For $\alpha = 2$ it is converted into a Wiener process.

In accordance with Eqs. (7)–(10), the solutions of the superdiffusion equation (11) belong to a class of strictly stable N-dimensional distributions[13,14] or, more accurately, form a subset of spherically symmetric distributions in this class which include the multidimensional Gaussian distribution ($\alpha = 2$). The most important characteristic of stable non-Gaussian distributions is that the absolute moments

$$\langle |x|^\mu \rangle \cong \int g^{(\alpha)}(x)|x|^\mu\, dx$$

are infinite for $\mu \geqslant \alpha$. Infinite dispersion implies that a different measure must be used for the width characteristic of the diffusion packet, for which it is convenient to take the radius $R_p(t)$ of a sphere containing the fixed probability p

$$\int_{|x| < R_p(t)} \rho^{(\alpha)}(x,t)\, dx = p.$$

Substituting Eq. (6) and changing the variable of integration, we obtain

$$\int_{|x| < R_p(t)(Dt)^{-1/\alpha}} g^{(\alpha)}(x)\, dx = p.$$

which implies that

$$R_p(t) \propto t^{1/\alpha}, \quad t \to \infty.$$

For $\alpha = 2$ we have a normal rate of expansion of a diffusion packet, whereas for $\alpha < 2$ its width increases more rapidly than in the normal case.

3. SPHERICALLY SYMMETRIC STABLE DISTRIBUTIONS

One-dimensional stable laws were analyzed in detail in Refs. 13 and 15. We shall discuss in greater detail some properties of the N-dimensional stable densities

$$g_N^{(\alpha)}(x) = (2\pi)^{-N} \int \exp\{-ikx - |k|^\alpha\}\, dk, \tag{12}$$

in terms of which the solutions of the superdiffusion equations are expressed in N-dimensional space. For the first three dimensions, allowing for the spherical symmetry of the characteristic functions, formula (12) has the following form:

$$g_1^{(\alpha)}(x)=\pi^{-1}\int_0^\infty e^{-s^\alpha}\cos(s|x|)\,ds,$$

$$g_2^{(\alpha)}(x)=(2\pi)^{-1}\int_0^\infty e^{-s^\alpha}J_0(s|x|)s\,ds,$$

$$g_3^{(\alpha)}(x)=(2\pi^2|x|)^{-1}\int_0^\infty e^{-s^\alpha}\sin(s|x|)s\,ds.$$

For arbitrary dimensions[13] we have

$$g_N^{(\alpha)}(x)=(2\pi)^{-N/2}\int_0^\infty e^{-s^\alpha}J_{N/2-1}(s|x|)\times(s|x|)^{1-N/2}s^{N-1}\,ds. \tag{13}$$

The functions $f_N(r)$ which determine the dependence of the spherically symmetric densities on the vector magnitude $r=|x|$

$$f_N(|x|)=g_N^{(\alpha)}(x)$$

are related by the differential equation

$$df_N/dr=-2\pi r f_{N+2}(r). \tag{14}$$

In addition to the integral representation (13), representations of the stable densities in the form of two series are also useful for computational and analytical purposes:

$$g_N^{(\alpha)}(x)=\frac{1}{\pi}(|x|\sqrt{\pi})^{-N}\times\sum_{n=1}^\infty(-1)^{n-1}\frac{\Gamma((n\alpha+N)/2)\Gamma(n\alpha/2+1)}{\Gamma(n+1)}\times\sin(\alpha n\pi/2)(|x|/2)^{-n\alpha},$$

$$g_N^{(\alpha)}(x)=\frac{2}{\alpha}(2\sqrt{\pi})^{-N}\times\sum_{n=0}^\infty(-1)^n\frac{\Gamma((2/\alpha)n+N/\alpha)}{\Gamma(n+N/2)\Gamma(n+1)}\left(\frac{|x|}{2}\right)^{2n}.$$

The first series converges for $\alpha\in(0,1)$ and is asymptotic for $\alpha\in(1,2)$, while the second converges in the range $\alpha\in[1,2]$ and is asymptotic for $\alpha\in(0,1)$.

Another integral representation for odd dimensions can be obtained from the symmetric one-dimensional density written in the form[13]

$$g_1^{(\alpha)}(x)=\frac{\alpha}{\pi|\alpha-1|}|x|^{1/(\alpha-1)}\int_0^{\pi/2}U_\alpha(\varphi)\times\exp\{-|x|^{\alpha/(\alpha-1)}U_\alpha(\varphi)\}\,d\varphi, \tag{15}$$

where

$$U_\alpha(\varphi)=\left(\frac{\sin(\alpha\varphi)}{\cos\varphi}\right)^{\alpha/(\alpha-1)}\frac{\cos((\alpha-1)\varphi)}{\cos\varphi},\quad \alpha\neq1,$$

is a series form of formula (14). Unlike formula (13), this representation does not contain oscillating functions in the integrand and is thus more convenient for numerical calculations (for $\alpha=1$ such a formula is not required, since the densities are expressed in an elementary form given below).

The densities $g_N^{(\alpha)}(x)$ are only expressed in terms of elementary or special functions in exceptional cases:

$$g_N^{(2)}(x)=(4\pi)^{-N/2}e^{-x^2/4}$$

— Gauss law (with a dispersion of 2),

$$g_N^{(1)}(x)=\Gamma((N+1)/2)[\pi(1+x^2)]^{-(N+1)/2}$$

— Cauchy law

$$g_N^{(2/3)}(x)=\frac{\sqrt{3}}{6\pi^{N/2}}\frac{\Gamma(N/2+1/3)\Gamma(N/2+2/3)}{\Gamma(5/6)\Gamma(7/6)}\times|x|^{-N}e^{2/(27x^2)}W_{-N/2,1/6}(4/(27x^2)),$$

where

$$W_{\nu,\mu}(z)=\frac{z^\nu e^{-z/2}}{\Gamma(\mu-\nu+1/2)}\times\int_0^\infty e^{-t}(1+t/z)^{\mu+\nu+1/2}t^{\mu-\nu-1/2}\,dt$$

is a Whittaker function. The density $g_N^{(1/2)}(x)$ for odd dimensions $N=2n+1$ may be expressed in the form

$$g_N^{(1/2)}(x)=\frac{1}{4\pi}\left(-\frac{1}{\pi}\right)^{(N-1)/2}\int_0^\infty t^{N-1}L_N(t|x|)\sin t\,dt,$$

where

$$L_N(\sqrt{s})=\frac{d^n}{ds^n}(\sqrt{s}+1/4)^{-3/2},\quad s>0.$$

For instance, we have

$$p_1^{(1/2)}(x)=\frac{1}{4\pi}\int_0^\infty(t|x|+1/4)^{-3/2}\sin t\,dt,$$

$$p_3^{(1/2)}(x)=\frac{3}{16\pi^2}|x|^{-1}\int_0^\infty(t|x|+1/4)^{-3/2}t\sin t\,dt,$$

and so on. We note however, that the formulas given above for $g_N^{(1/2)}$ contain an alternating-sign function in the integrand and consequently have no particular advantages over the formulas of the inverse Fourier transformation (12) and (13).

We shall discuss another distribution which describes a random electric or gravitational field intensity created by a Poisson ensemble of point sources (Holtsmark distribution):

$$g_3^{(3/2)}(x)=\frac{1}{2\pi^2|x|^3}\int_0^\infty\exp\{-(z/|x|)^{3/2}\}z\sin z\,dz$$

$$=\frac{3}{2\pi^2}\int_0^{\pi/2}[3|x|^3U_{3/2}^2(\varphi)-2U_{3/2}(\varphi)]\times\exp\{-|x|^3U_{3/2}(\varphi)\}\,d\varphi,$$

$$\int_{|x|>r} g_3^{(3/2)}(x)\,dx=\frac{2}{\pi}\int_0^{\pi/2}(1+3U_{3/2}(\varphi)r^3)e^{-U_{3/2}(\varphi)r^3}\,d\varphi.$$

The distribution of peculiar velocities of galaxies in the universe is associated with this distribution which can then be used to estimate the most important cosmological parameter, viz., the average density of matter.[16]

Values of the densities $g_3^{(\alpha)}(x)$ and their graphical representation can be found in Refs. 17 and 18.

4. TRANSPORT EQUATIONS

It was noted in Sec. 2 that the width of a diffusion packet described by a Lévy process increases as $t^{1/\alpha}$. For $\alpha<1$ it increases more rapidly than in the ballistic regime (i.e., free motion of particles with a bounded maximum velocity). This clearly nonphysical result is attributable to the self-similarity of the Lévy process, in which the concept of the velocity of free particle motion has no place. This situation is not confined to $\alpha<1$, but the effect is observed more clearly in this limit. For $\alpha>1$ as far as the limiting value $\alpha=2$, the self-similarity effect is manifested in the way a distribution $\rho(x,t)$ at a time arbitrarily close to the initial one (when the particle was located at the origin) is nonzero throughout all space (this ''defect'' was noted by Einstein in the theory of normal diffusion).

This defect can be eliminated by converting from the Wiener model to the random walk model of a particle with a finite velocity of free motion v.

We shall consider the following model. At zero time $t=0$ the particle is located at the origin $x=0$ and resides there for a random time τ_0, after which it is displaced by a random vector ξ_1 at velocity v and again resides in a rest state for a random time τ_1, and the process then repeats. All random variables $\tau_0,\ \tau_1,\ \xi_1,\ \tau_2,\ \xi_2,\ \ldots$ are mutually independent and the times τ_i have the same exponential probability density

$$q(t)=\mu e^{-\mu t},\quad \mu>0.$$

The N-dimensional vectors ξ_i are also distributed similarly.

Instead of a single particle, it is convenient to consider a set of independent trajectories and talk of the density $\rho(x,t)$ as the particle number density.

Hence in this particular case the particle density $\rho(x,t)$ consists of two components $\rho_0(x,t)$ and $\rho_v(x,t)$, which refer to particles in the rest state and in motion, respectively:

$$\rho(x,t)=\rho_0(x,t)+\rho_v(x,t). \tag{16}$$

Over the time dt the density of particles in traps changes by the amount

$$d\rho_0(x,t)=\rho_0(x,t+dt)-\rho_0(x,t),$$

which consists of two components. The first, negative, component is attributed to particles leaving the traps

$$[d\rho_0(x,t)]_-=-\mu\rho_0(x,t)dt,$$

while the second, positive, component describes the entrance of moving particles into traps:

$$[d\rho_0(x,t)]_+=\int dx'\,p(x')\mu\rho_0(x-x',t-|x'|/v)\,dt.$$

As a result we obtain the following equation for $\rho_0(x,t)$:

$$\frac{\partial\rho_0}{\partial t}=-\mu\rho_0+\mu\int dx'\,p(x')\rho_0\left(x-x',t-\frac{|x'|}{v}\right). \tag{17}$$

Having noted that $p(x)dx$ is the probability that a particle leaving the origin will undergo the first collision in the volume element $dx=dSd|x|$, we denote by $P(x)dS$ the probability that the particle intersects the elementary area dS of a sphere of radius $|x|$ without interacting over the path $|x|$. The contribution to the density ρ_v of this particle is $(1/v)P(x)\delta(t-|x|/v)$. Replacing x by x' and applying this result to all particles leaving the traps and reaching the vicinity of the point x at time t, we obtain

$$\rho_v(x,t)=\frac{1}{v}\int dx'\int dt'\,P(x')\delta\left(t'-\frac{|x'|}{v}\right)\mu\rho_0(x-x',t-t')$$
$$=\frac{1}{v}\int dx'P(x')\mu\rho_0(x-x',t-t'). \tag{18}$$

In three-dimensional space with an exponential range distribution the system of equations (17) and (18) describes the nonsteady-state transport of neutrons allowing for a delay and, apart from a few details (the absence of absorption and scattering processes, and constant velocity), adequately describes Eqs. (1.13) and (1.14) from Ref. 19. For $\mu\to0$ this system yields the time-dependent single-velocity transport equation with isotropic scattering widely used in neutron physics.[19,20] In view of this we shall describe Eqs. (18) and (19) as transport equations.

In the limit $v=\infty$, only one term is retained in the sum (16):

$$\rho(x,t)=\rho_0(x,t),$$

which satisfies the Kolmogorov equation

$$\frac{\partial\rho}{\partial t}=-\mu\rho+\mu\int dx'\,p(x')\rho(x-x',t), \tag{19}$$

describing a generalized Poisson process.[21]

Returning to a probabilistic interpretation of these equations, we shall consider a common initial condition for them:

$$\rho(x,0)=\rho_0(x,0)=\delta(x).$$

5. RELATION BETWEEN SOLUTIONS OF THE KOLMOGOROV EQUATION AND THE SUPERDIFFUSION EQUATION

We shall transform the Kolmogorov equation (19) to give an equation for the characteristic function $\hat\rho(k,t)$ of the distribution $\rho(x,t)$:

$$\frac{\partial\hat\rho}{\partial t}=-\mu[1-\hat p(k)]\hat\rho(k,t),\quad \hat\rho(k,0)=1, \tag{20}$$

where $\hat p(k)$ is the Fourier component of the probability density of the transition $p(x)$. The solution of Eq. (20) has the form

$$\hat{\rho}(k,t)=\exp\{-[1-\hat{p}(k)]\mu t\}, \tag{21}$$

and its asymptotic behavior for $t\to\infty$ is determined by the behavior of $\hat{p}(k)$ for small $|k|$.

If the second moment of the distribution $p(x)$ is

$$\int p(x)x^2\,dx=\overline{\xi^2}<\infty, \tag{22}$$

then

$$1-\hat{p}(k)\sim(\overline{\xi^2}/2)k^2, \quad |k|\to 0,$$

and the characteristic function (21) for large t has the form

$$\hat{\rho}(k,t)\sim\hat{\rho}_{\rm as}(k,t)=\exp\{-(\mu t\overline{\xi^2}/2)k^2\}. \tag{23}$$

Since

$$\frac{\partial\hat{\rho}_{\rm as}(k,t)}{\partial t}=-\frac{\mu\overline{\xi^2}}{2}k^2\hat{\rho}_{\rm as}(k,t),$$

the density corresponding to this asymptotic form satisfies the ordinary diffusion equation

$$\frac{\partial\rho_{\rm as}}{\partial t}=D\Delta\rho_{\rm as}(x,t)$$

with the diffusion coefficient

$$D=\mu\overline{\xi^2}/2$$

and the initial condition

$$\rho_{\rm as}(x,0)=\delta(x).$$

We have focused on this generally trivial fact in order to emphasize that the asymptotic form of the solution of Eq. (19) subject to the condition (22) is an exact solution of the ordinary diffusion equation. When condition (22) is not satisfied but the following condition is,

$$\int_{|x|>R}p(x)\,dx\sim AR^{-\alpha}, \quad R\to\infty, \tag{24}$$

the asymptotic form of the solution of Eq. (19) is the same as the exact solution of the superdiffusion equation, which is easily confirmed bearing in mind that

$$1-\hat{p}(k)\sim A'|k|^{\alpha}, \quad k\to 0,$$

and deriving instead of Eq. (23)

$$\hat{\rho}(k,t)\sim\hat{\rho}_{\rm as}(k,t)=\exp\{-\mu tA'|k|^{\alpha}\}, \quad t\to\infty.$$

In accordance with Eqs. (6) and (10), we have

$$\rho_{\rm as}(k,t)=\rho^{(\alpha)}(x,t)=(Dt)^{-N/\alpha}g^{(\alpha)}(x(Dt)^{-1/\alpha}), \tag{25}$$

where

$$D=\mu A'.$$

In other words, the superdiffusion equation (11) describes the asymptotic behavior of the distribution density of a particle wandering at infinite velocity in a medium containing traps for which the residence time τ has an exponential distribution and the range (jump) distribution density has a power tail of the form $r^{-\alpha-1}$.

The result (25) can also be obtained more simply on the basis of the properties of stable laws. In the same way that, provided that the second moment (22) is finite, the normalized sum

$$S_n=(Bn)^{-1/2}\sum_{i=1}^{n}\xi_i, \quad B=\overline{\xi^2}$$

of n independent random vectors ξ_i for $n\to\infty$ has an N-dimensional Gaussian distribution (central limit theorem), if condition (24) is satisfied the normalized sum

$$S_n=(Bn)^{-1/\alpha}\sum_{i=1}^{n}\xi_i, \quad B=B(\alpha,A), \tag{26}$$

for large n is described by a stable distribution with the exponent α (generalized limit theorem). The distribution of the number of terms over time t is given by the Poisson law with the average $\overline{n}=\mu t$ and the relative fluctuations $(\mu t)^{-1/2}$. In view of this, for $\mu t\to\infty$ in formula (26), n may be replaced by μt. Bearing in mind that as a result of this substitution $\Sigma_{i=1}^{n}\xi_i$ yields the random vector $\xi(t)$ of the particle position at time t, we obtain

$$\xi(t)=(Dt)^{1/\alpha}\zeta, \quad D=B\mu, \tag{27}$$

where ζ is a random vector with a symmetric stable density $g^{(\alpha)}(x)$, from which formula (25) follows.

6. EFFECT OF FINITE VELOCITY

In order to allow for the influence of the finite velocity of free motion of a wandering particle on the asymptotic form of the distribution $\rho(x,t)$, we need to return to Eqs. (17) and (18) and make an analysis similar to that made in Sec. 5 for the Kolmogorov equation. We performed such an analysis, but because of the involved nature of the mathematical calculations we shall confine ourselves here to an elementary derivation which yields the same result as a more rigorous analysis.

For a finite velocity v the vector sum

$$S_n=(Bn)^{-1/\alpha}\sum_{i=1}^{n}\xi_i,$$

corresponds to the random time

$$\theta=\sum_{i=1}^{n}\left(\tau_i+\frac{|\xi_i|}{v}\right).$$

In the case $\alpha>1$ the mathematical expectation $|\xi_i|$ is finite, and for $n\to\infty$, since the numbers are large, we can assume

$$\theta\approx t=n(1/\mu+a/v), \quad a=\overline{|\xi_i|}. \tag{28}$$

Then by finding n

$$n=(1+\mu a/v)^{-1}\mu t$$

and introducing the notation

$$t_v=(1+\mu a/v)^{-1}t,$$

we arrive at the same result as in Sec. 5, except that t_v appears instead of t:

$$\hat{\rho}_{as}(x,t)=(Dt_v)^{-N/\alpha}g^{(\alpha)}(x(Dt_v)^{-1/\alpha}),\quad \alpha>1. \qquad (29)$$

Physically, this result is quite understandable: the presence of a finite velocity of free motion slows the expansion of a diffusion packet compared with the case $v=\infty$. Replacing t by the shorter time t_v immediately allows for this slowing (in the asymptotic sense).

Since the diffusion coefficient and the time appear as a product in the asymptotic density, the result (29) may be rewritten in a different form:

$$\rho_{as}(x,t)=(D_v t)^{-N/\alpha}g^{(\alpha)}(x(D_v t)^{-1/\alpha}),\quad \alpha>1,$$

where

$$D_v=(1+\mu a/v)^{-1}D.$$

This allows us to write a fractional differential equation for the asymptotic density of a particle diffusing with a finite velocity of free motion:

$$\frac{\partial \rho_{as}}{\partial t}=-D_v(-\Delta)^{\alpha/2}\rho_{as}(x,t).$$

Thus, allowance for a finite velocity merely alters (reduces) the diffusion coefficient in the equation with a fractional Laplacian, which preserves the form of the asymptotic distribution described by a spherically symmetric stable law. However, this conclusion only holds for $\alpha>1$, which was used for the substitution (28). For $\alpha<1$ the situation is completely different: in cases of finite velocity the asymptotic distribution cannot be reduced to a solution with $v=\infty$ by any linear transformation.

This last factor is easily understood through the following reasoning. It was noted in Sec. 2 that the width (''radius'') of a diffusion packet increases with time proportionately as $t^{1/\alpha}$. The presence of a finite velocity causes the distribution density to vanish outside a sphere of radius $R=vt$. Hence for $\alpha>1$ the influence of the last (kinematic) constraint becomes weaker with time because the packet radius increases more slowly than $R=vt$. However, for $\alpha<1$ the distribution given by the diffusion equation (11) spreads more rapidly than $R=vt$ and the kinematic constraint becomes a dominant factor in the formation of the asymptotic distribution. Being compressed by a sphere of radius $R=vt$, this distribution is completely different to the stable distribution. This probably means that Eq. (12) with a Laplacian of degree $\alpha/2$, where $\alpha<1$, cannot generally be applied to describe real diffusion processes.

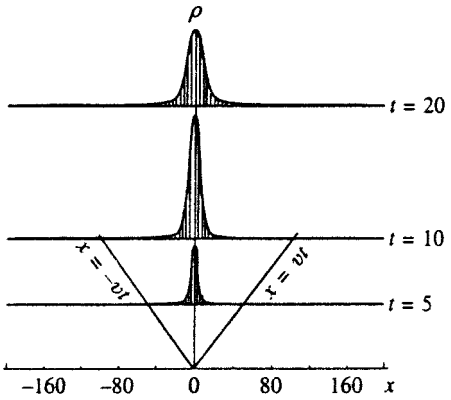

FIG. 1. The distribution $\rho(x,t)$ in cases of one-dimensional superdiffusion for $\alpha=3/2$, $\mu=1$. The histograms give the results of Monte Carlo simulations (10^4 trajectories, $v=10$), which give the solution of the transport equations (17) and (18). The solid curves give the solutions of the hyperdiffusion equation (11) using D_v instead of D.

The results of a Monte Carlo simulation of the one-dimensional random walk of a particle compared with the solutions of the superdiffusion equation plotted in Figs. 1–3 confirm the conclusions reached above: it can be seen that

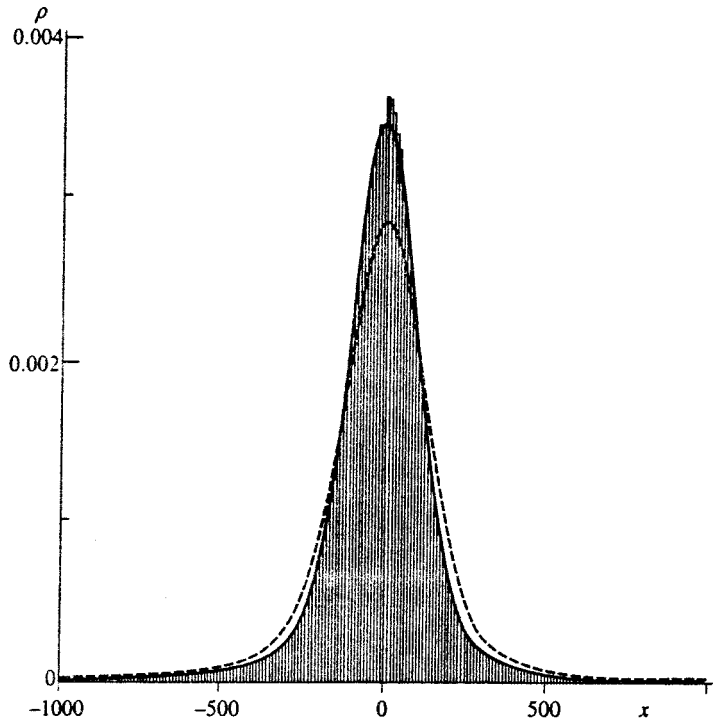

FIG. 2. Effect of finite velocity ($\alpha=3/2$, $\mu=1$, and $t=10^3$. The histogram gives the results of a Monte Carlo simulation (2×10^5 trajectories, $v=5$) of the solution of the transport equation; the dashed curve gives the solution of Eq. (11) with the diffusion coefficient D ($v=\infty$), the solid curve gives the solution with the coefficient D_v.

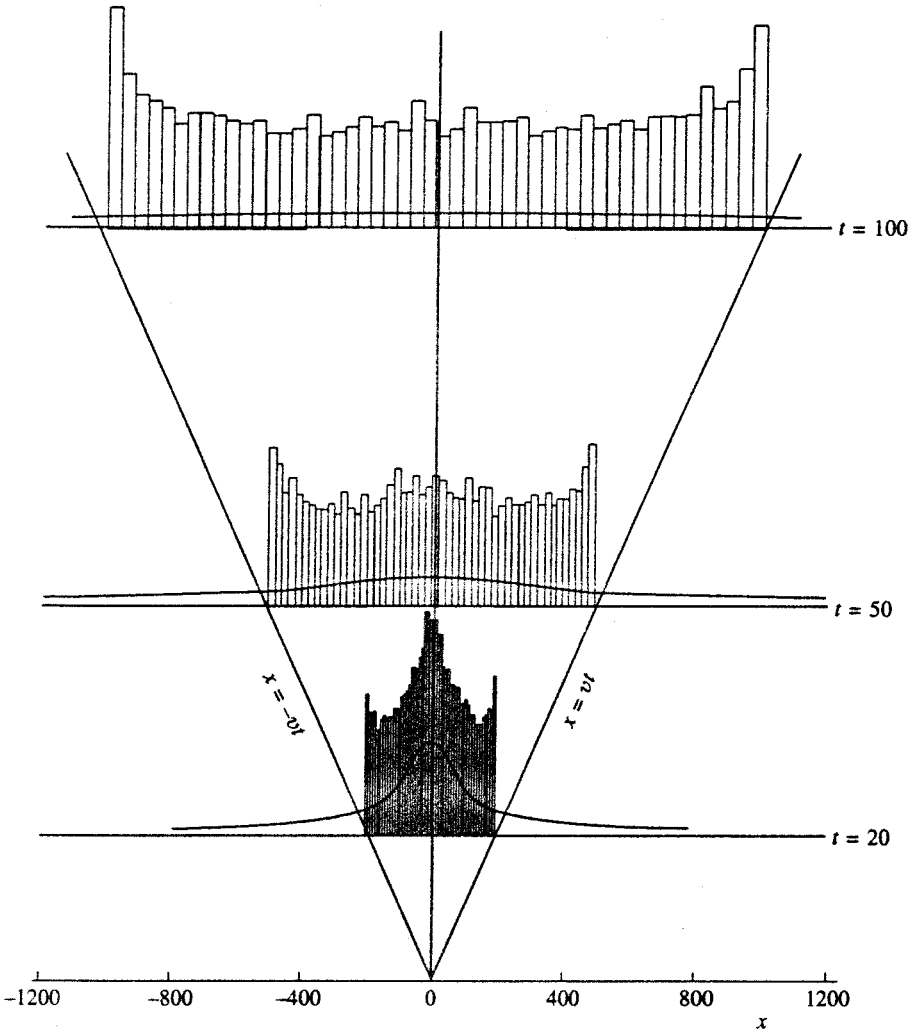

FIG. 3. The distribution $\rho(x,t)$ for one-dimensional superdiffusion for $\alpha=1/2$ and $\mu=1$. The histograms give the results of the Monte Carlo simulation (10^4 trajectories, $v=10$), the solid curves give the solution of Eq. (11).

for $\alpha=3/2$ replacing D by D_v ensures asymptotic agreement between the solutions of the superdiffusion and transport equations, whereas for $\alpha=1/2$ the solutions of these equations are completely different.

7. CONCLUSIONS

1. The superdiffusion equation (11) describes the asymptotic behavior of a generalized Poisson process with instantaneous (discontinuous) increments whose absolute value is distributed with the density $p(r)\propto r^{-\alpha-1}$, $0<\alpha<2$.

2. The solutions of Eq. (11) belong to a class of spherically symmetric stable distributions $g^{(\alpha)}(x)$ whose properties have been described.

3. For $\alpha\in[1,2]$ the superdiffusion equation describes the asymptotic behavior of the distribution of a wandering particle having the finite velocity v of free motion (provided that D is replaced by $D_v=(1+\mu a/v)^{-1}D$, where a is the mean range and $1/\mu$ is the average residence time in the trap).

4. For $\alpha<1$ a superdiffusion packet spreads more rapidly in space than a packet of freely moving particles, and the solutions of the superdiffusion and transport equations have completely different asymptotic forms.

This last observation may be a sufficient basis for concluding that the superdiffusion equation cannot be used to describe real physical processes for characteristic exponents $\alpha<1$.

[a)]E-mail: uchaikin@themp.univ.simbirsk.su

[1] J. -P. Bouchard and A. Georges, Phys. Rep. **195**, 127 (1990).
[2] M. B. Isichenko, Rev. Mod. Phys. **64**, 961 (1992).
[3] R. R. Nigmatullin, Phys. Status Solidi B **123**, 739 (1984).
[4] R. Metzler, W. G. Glockle, and T. F. Nonnenmacher, Physica A **211**, 13 (1994).
[5] H. C. Fogedby, Phys. Rev. Lett. **73**, 2517 (1994).
[6] K. V. Chukbar, Zh. Éksp. Teor. Fiz. **108**, 1875 (1995) [JETP **81**, 1025 (1995)].
[7] A. S. Monin, Dokl. Akad. Nauk SSSR **105**, 256 (1955).
[8] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics*, Vol. 2 (MIT Press, Cambridge, Mass., 1975) [Russ. original, Vol. 2, Nauka, Moscow, 1967, p. 509).
[9] A. Compte, Phys. Rev. E **53**, 4191 (1996).
[10] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives—Theory and Applications* (Gordon and Breach, New York, 1993) [Russ. original, Nauka i Tekhnika, Minsk, 1987, p. 357).
[11] P. Lévy *Stochastic Processes and Brownian Motion* [Russ. transl., Nauka, Moscow, 1972, p. 164].
[12] E. W. Montroll and B. J. West, in *Fluctuation Phenomena*, edited by E. W. Montroll and J. L. Lebowitz (North-Holland, Amsterdam, 1979), p. 61.
[13] V. M. Zolotarev, *One-Dimensional Stable Distributions*, [in Russian], Mir, Moscow (1983).

[14] V. M. Zolotarev, in *Contributions to Probability*, edited by J. Gam and V. K. Rohatgi (Academic Press, New York, 1981), p. 283.
[15] B. V. Gnedenko and A. N. Kolmogorov, *Limiting Distributions for Sums of Independent Random Quantities* [in Russian], GITTL, Moscow (1949).
[16] M. Rowan-Robinson, A. Lawrence, W. Saunders *et al.*, Mon. Not. R. Astron. Soc. **247**, 1 (1990).
[17] V. V. Uchaikin and G. G Gusarev, *Proceedings of the Third Symposium on the Renormalization Group*, OIYaI, Dubna, 1997, p. 417.
[18] V. V. Uchaikin and G. G. Gusarev, J. Math. Phys. **38**, 2453 (1997).
[19] S. B. Shikhov, *Problems in the Mathematical Theory of Reactors* [in Russian], Atomizdat, Moscow (1973).
[20] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967) [Russ. transl., Mir, Moscow, 1972].
[21] W. Feller, *Introduction to Probability Theory and Its Applications*, 3rd. ed. (Wiley, New York, 1967) [Russ. transl., later ed., Vol. 2, Nauka, Moscow, 1984].

Translated by R. M. Durham

Emission of radiation from a long Josephson junction in a thin film

A. S. Malishevskiĭ,*) V. P. Silin, and S. A. Uryupin

P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 14 July 1998)
Zh. Éksp. Teor. Fiz. **115**, 1426–1449 (April 1999)

Space–time nonlocal electrodynamic equations are derived for nonlinear vortex states of a Josephson junction in a film of thickness much smaller than the London penetration depth. The spectrum and damping of generalized Swihart waves propagating in such a junction are analyzed. The radiation damping constant associated with the possible emission of electromagnetic radiation is determined in the range of Swihart wave phase velocities exceeding the speed of light. The emission of radiation from nonlinear states having dimensions greater than the distance traversed by light in vacuum during the characteristic time of variation of the phase difference is investigated. It is shown that the flux density of radiation emitted by such states is localized in a plane orthogonal to the axis of the tunnel junction and depends weakly on the angle of observation in this plane.
[S1063-7761(99)01904-6]

1. INTRODUCTION

The radiative properties of Josephson junctions has attracted the attention of specialists for many years (see, e.g., Refs. 1–3). Many theoretical and experimental papers (see the survey in Barone and Paterno,[3] Sec. 11.4) have been devoted to the investigation of microwave radiation from Josephson junctions having dimensions much smaller than the wavelength of the radiation itself. Another avenue of research is directed toward the properties of long Josephson junctions, along which it is possible for both Swihart waves and solitons to propagate. Long junctions afford the possibility of studying such phenomena as the interaction of a Swihart wave with a surface electromagnetic wave[4] and the emission of electromagnetic waves by moving vortices.[5–9]

Another property of long junctions is the possibility of the emission of Swihart waves and the existence of nonlinear vortex states. A quantitative description of the emission of radiation from a waveguide sandwich has been described[10] on the basis of a systematic analysis of electromagnetic fields in superconducting electrodes and vacuum. According to Ref. 10, the emission of electromagnetic waves leads to further damping of the Swihart waves, whose phase velocities are higher than the speed of light. The emission phenomenon is most conspicuous when the thickness of the sandwich electrodes is smaller than the London penetration depth, and the electromagnetic field easily penetrates from the Josephson junction into vacuum.

The emission of electromagnetic waves can be expected to be even simpler for a Josephson junction in a thin film whose thickness D is much smaller than the London depth λ. Another feature of Josephson thin-film electrodynamics is the large penetration depth of the field into the film, $\lambda_e=\lambda^2/D\gg\lambda$ (Ref. 11), which makes it critical to utilize the nonlocal coupling of the magnetic field with the phase difference of the wave functions on opposite sides of a tunnel junction.[12]

Bearing these special features in mind, in Sec. 2 we derive an equation for the phase difference, taking into account both the spatially nonlocal coupling effect and the possibility of emitting electromagnetic waves. We obtain a new space–time nonlocal integrodifferential equation for the phase difference, which goes over to the standard equation[12–14] when time dispersion is disregarded. Using this type of nonlocal equation for the phase difference, in Sec. 3 we find a dispersion law for generalized Swihart waves. We show that in the range of phase velocities lower than the speed of light the dispersion law differs from the previously established law[14] when interaction between a surface electromagnetic wave and a Swihart wave is significant.

The difference is even more pronounced in the limit of superluminous phase velocity, when not only is the real part of the Swihart wave frequency renormalized, but additional damping is also induced by the emission of waves. The radiative damping constant is determined, and it is shown that in a low-conductivity tunnel junction it can exceed the damping constant associated with Ohmic losses.

In Sec. 4 we determine the directivity pattern of the radiation emitted by the Swihart wave. If the ratio $v=\omega/ck_z$ of the Swihart phase velocity to the speed of light is close to unity, $0<v^2-1\ll 1$, waves are emitted along the tunnel junction. But in the limit $v\gg 1$ the radiation flux $\mathbf{S}$ is mainly concentrated in a plane orthogonal to the axis of the junction.

We find the explicit dependence of the flux density on the angle of observation in this same plane.

Section 5 is devoted to the emission of large-scale nonlinear distributions of the phase difference with characteristic dimensions much smaller than the distances traversed by light in vacuum during the characteristic time of variation of the phase difference. Such distributions are described in the first approximation by the same functions as the nonlinear oscillations of a mathematical pendulum. We show that the radiation flux emitted by large-scale nonlinear structures is concentrated mainly in a plane orthogonal to the axis of the

tunnel junction and is constant over a wide range of angles of observation in this plane. The spectral composition of the radiation depends on which solution of the mathematical pendulum problem describes the nonlinear state. The spectrum of the radiation contains odd harmonics of the fundamental for the solution corresponding to finite-amplitude oscillations, and it contains even harmonics for the solution corresponding to a rotating pendulum. If the nonlinear state is described by a solution of the 2π-kink type, a continuous spectrum of frequencies lower than or of the same order as the Josephson frequency is radiated.

2. BASIC EQUATIONS

We consider a Josephson junction in a superconducting thin film symmetrical about the plane $y=0$. We assume that the thickness of the film D along the y axis is much smaller than the London penetration depth λ. We also assume that the tunneling of Cooper pairs takes place through a nonsuperconducting layer of thickness $2d$ oriented symmetrically about the axis $x=y=0$. To characterize the electrodynamics of this junction, we use the equations for the vector potential $\mathbf{A}(\mathbf{r},t)$ and the scalar potential $V(\mathbf{r},t)$ in the form

$$\operatorname{div}\mathbf{A}(\mathbf{r},t)=0, \tag{2.1}$$

$$\nabla^2\mathbf{A}(\mathbf{r},t)-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\mathbf{A}(\mathbf{r},t)-\frac{1}{c}\frac{\partial}{\partial t}\operatorname{grad}V(\mathbf{r},t) =-\frac{4\pi}{c}\delta(y)D\mathbf{j}(\mathbf{r}_\parallel,t), \tag{2.2}$$

$$\nabla^2V(\mathbf{r},t)=0, \tag{2.3}$$

where c is the speed of light, $\delta(y)$ is the delta function, $\mathbf{j}(\mathbf{r}_\parallel,t)$ is the current density in the film, $\mathbf{r}_\parallel=(x,0,z)$,

$$\mathbf{j}(\mathbf{r}_\parallel,t)=-\frac{c}{4\pi\lambda^2}\left[\frac{\phi_0}{2\pi}\operatorname{grad}\Phi(\mathbf{r}_\parallel,t)+\mathbf{A}(y=0,\mathbf{r}_\parallel,t)\right], \tag{2.4}$$

$\phi_0=\pi\hbar c/|e|$ is the quantum of magnetic flux, $\hbar$ is Planck's constant, e is the electron charge,

$$\Phi(\mathbf{r}_\parallel,t)=\eta(-x)\Phi_1(\mathbf{r}_\parallel,t)+\eta(x)\Phi_2(\mathbf{r}_\parallel,t), \tag{2.5}$$

$\eta(x)=1$ $(x>0)$ and $\eta(x)=0$ $(x<0)$ denotes the Heaviside unit step function, and $\Phi_1(\mathbf{r}_\parallel,t)$ and $\Phi_2(\mathbf{r}_\parallel,t)$ are the phases of the wave functions of Cooper pairs on opposite sides of the tunnel junction. The displacement current, which is small in comparison with the superconducting current in the film, is omitted in Eq. (2.4). Moreover, Eq. (2.5) disregards the finite thickness of the tunnel junction, and the expression for the superconducting current density in the film [see Eqs. (2.2) and (2.4)] is written in the form of the delta function $\delta(y)$. Equations (2.2), (2.4), and (2.5) are therefore suitable for describing electromagnetic structures having spatial scales that exceed both the width of the tunnel junction $2d$ and the thickness of the film D.

Owing to the gauge invariance of the theory, the scalar potential $V(y=0,\mathbf{r}_\parallel,t)$ in the superconducting film is related to the phase of the wave function by the equation (see, e.g., Ref. 15)

$$V(y=0,\mathbf{r}_\parallel,t)=\frac{\phi_0}{2\pi c}\frac{\partial}{\partial t}\Phi(\mathbf{r}_\parallel,t). \tag{2.6}$$

Equations (2.1)–(2.6) can be used to relate the electromagnetic potential to the phase difference at the junction:

$$\varphi(z,t)=\Phi_1(0,z,t)-\Phi_2(0,z,t). \tag{2.7}$$

It is evident from the symmetry of the problem and Eqs. (2.1)–(2.4) that the potentials $A_x(\mathbf{r},t)$, $A_z(\mathbf{r},t)$, and $V(\mathbf{r},t)$ are even, and $A_y(\mathbf{r},t)$ is an odd function of the argument y. It is therefore sufficient to find a solution of Eqs. (2.1)–(2.4) for $y\geqslant 0$ and to use the boundary conditions

$$\frac{\partial}{\partial y}\mathbf{A}_\parallel(y,\mathbf{r}_\parallel,t)\big|_{y=+0}=\frac{1}{2\lambda_e}\left[\frac{\phi_0}{2\pi}\frac{\partial}{\partial\mathbf{r}_\parallel}\Phi(\mathbf{r}_\parallel,t)+\mathbf{A}_\parallel(y=0,\mathbf{r}_\parallel,t)\right], \tag{2.8}$$

$$A_y(y=0,\mathbf{r}_\parallel,t)=0, \tag{2.9}$$

where the vector $\mathbf{A}_\parallel$ has only the components A_x and A_z. Equation (2.8) follows from Eq. (2.2) when the continuity of the potentials $\mathbf{A}_\parallel(\mathbf{r},t)$ and $V(\mathbf{r},t)$ at $y=0$ and the oddness of the derivatives $\partial\mathbf{A}_\parallel/\partial y$ is taken into account. Condition (2.9) follows from the continuity and oddness of the function A_y and implies that the current does not have a component j_y given by (2.4) normal to the surface of the film. We write the solution of Eqs. (2.1)–(2.4) in the domain $y\geqslant 0$, subject to the boundary conditions (2.8) and (2.9), in the form (see Appendix A)

$$\begin{Bmatrix}\mathbf{A}(\mathbf{r},t)\\ V(\mathbf{r},t)\end{Bmatrix}=\frac{\phi_0}{(2\pi)^4}\int_{-\infty}^{\infty}d\omega\int d\mathbf{k}\int_{-\infty}^{\infty}dt'\times\int_{-\infty}^{\infty}dz'\,\varphi(z',t')\exp[i\omega(t'-t)+i\mathbf{k}\cdot\mathbf{r}_\parallel-ik_zz']\frac{1}{G}\begin{Bmatrix}\mathbf{g}\\ g_v\end{Bmatrix}. \tag{2.10}$$

In Eq. (2.10) we have introduced the functions

$$g_x=\frac{k_x^2}{k}\exp(-ky)-\frac{G+2\lambda_ek_x^2}{1+2\lambda_e\psi}\exp(-\psi y), \tag{2.11}$$

$$g_y=ik_x[\exp(-ky)-\exp(-\psi y)], \tag{2.12}$$

$$g_z=\frac{k_xk_z}{k}\left[\exp(-ky)-\frac{2\lambda_ek}{1+2\lambda_e\psi}\exp(-\psi y)\right], \tag{2.13}$$

$$g_v=\frac{\omega}{c}\frac{k_x}{k}\exp(-ky), \tag{2.14}$$

and G and ψ are given by the expressions

$$G=\psi-2\lambda_e\omega^2/c^2, \tag{2.15}$$

$$\psi=\sqrt{k^2-\omega^2/c^2}\;\eta(c^2k^2-\omega^2)-i\sqrt{\omega^2/c^2-k^2}\;\eta(\omega^2-c^2k^2)\operatorname{sgn}\omega. \tag{2.16}$$

Augmented with the notation (2.11)–(2.16), Eq. (2.10) establishes coupling, nonlocal in time and space, between the electromagnetic potentials and the phase difference of the wave functions $\varphi(z,t)$. The integrand in (2.10) has a singularity at $G=0$, where

$$\frac{\omega^2}{c^2k^2}=\frac{1}{8\lambda_e^2k^2}(\sqrt{1+16\lambda_e^2k^2}-1)\leqslant 1. \tag{2.17}$$

Equation (2.17) describes the dispersion law of a surface electromagnetic wave for which the components of the vector potential A_x and A_z are even, and A_y is an odd function of the variable y. The singularity at $G=0$ is correctly bypassed if the energy losses of the surface electromagnetic wave in the superconducting film are taken into account.

The above solution of Eqs. (2.1)–(2.6) with the boundary conditions (2.8) and (2.9) enables us to relate the vector and scalar potentials in vacuum and in the superconducting film to the phase jump between superconductors separated by a contact, whose physical properties have not yet been taken into account. The solution (2.10)–(2.17) is obtained from Eqs. (2.1)–(2.4), (2.6), which are applicable outside the contact, but this solution, in turn, gives the form of the solution only as a result of the presence of the phase jump in transition across the contact [see Eqs. (2.5)–(2.7)].

This problem has to be solved in order to obtain an equation for the phase difference φ at the film-separating transition layer, which corresponds to the phase jump of the superconductor wave functions. Such an equation can be derived by the approach of Refs. 12 and 13, according to which Eqs. (2.10)–(2.17) are used to calculate the x-component of the supercurrent density at the contact [cf. (2.4)]

$$j_x(x=0,z,t)=-\frac{c}{4\pi\lambda^2}\left[\frac{\phi_0}{2\pi}\frac{\partial}{\partial x}\Phi(\mathbf{r}_\parallel,t)\big|_{x=0}+A_x(x=0,y=0,z,t)\right]$$

in terms of the phase difference at the contact. Here, according to Eqs. (2.6) and (2.10), we have

$$j_x(x=0,z,t)=\frac{c}{4\pi\lambda^2}\frac{\phi_0}{(2\pi)^4}\int_{-\infty}^{\infty}d\omega\int d\mathbf{k}\int_{-\infty}^{\infty}dt'\int_{-\infty}^{\infty}dz'\times\varphi(z',t')\exp[ik_z(z-z')-i\omega(t-t')]\times\frac{1}{G}\frac{G+2\lambda_ek_x^2}{1+2\lambda_e\psi}.$$

On the other hand, according to Refs. 12 and 13, the supercurrent through the Josephson contact is equal to the sum[1–3] of the Josephson current

$$j_c\sin\varphi(z,t), \tag{2.18}$$

the conduction current

$$\sigma E_x(x=0,y=0,z,t)=\frac{\hbar\sigma}{4|e|d}\frac{\partial}{\partial t}\varphi(z,t) \tag{2.19}$$

and the displacement current

$$\frac{\varepsilon}{4\pi}\frac{\partial}{\partial t}E_x(x=0,y=0,z,t)=\frac{\hbar\varepsilon}{16\pi|e|d}\frac{\partial^2}{\partial t^2}\varphi(z,t) \tag{2.20}$$

where j_c is the critical Josephson current density, and σ and ε are the conductivity and dielectric permittivity of the tunnel junction, respectively.

Thus, the sum of the currents (2.18)–(2.20) must be set equal to the x-component of the current (2.4) at $x=0$, so that

$$\sin\varphi(z,t)+\frac{\beta}{\omega_j^2}\frac{\partial}{\partial t}\varphi(z,t)+\frac{1}{\omega_j^2}\frac{\partial^2}{\partial t^2}\varphi(z,t)=\frac{j_x(x=0,z,t)}{j_c}, \tag{2.21}$$

where $\beta=4\pi\sigma/\varepsilon$, and $\omega_j=4\pi(dcj_c/\varepsilon\phi_0)^2$ is the Josephson frequency. Using the explicit expression for $j_x(x=0,z,t)$, we arrive at an integrodifferential equation for the phase difference:

$$\sin\varphi(z,t)+\frac{\beta}{\omega_j^2}\frac{\partial}{\partial t}\varphi(z,t)+\frac{1}{\omega_j^2}\frac{\partial^2}{\partial t^2}\varphi(z,t)=-4l\lambda_e\int_{-\infty}^{\infty}dz'\int_{-\infty}^{\infty}dt'\varphi(z',t')Q(z-z',t-t'), \tag{2.22}$$

where $l=\lambda_j^2/\lambda$, $\lambda_j=(c\phi_0/\lambda j_c)^{1/2}/4\pi$ is the Josephson wavelength, and the kernel $Q(z,t)$ of the space–time nonlocal coupling has the form

$$Q(z,t)=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\int_{-\infty}^{\infty}\frac{dk_z}{2\pi}\exp(ik_zz-i\omega t)Q(k_z,\omega), \tag{2.23}$$

$$Q(k_z,\omega)=\frac{dk_x}{2\pi}\frac{1}{k^2}\left[\frac{\psi k_z^2}{1+2\psi\lambda_e}-\frac{\omega^2k_x^2}{\psi c^2-2\lambda_e\omega^2}\right]\equiv Q'(k_z,\omega)+iQ''(k_z,\omega). \tag{2.24}$$

Of course, there is an alternative approach (see, e.g., Ref. 1) utilizing the Maxwell equation at the contact

$$\operatorname{curl}\mathbf{H}=\frac{4\pi}{c}\mathbf{j}+\frac{\varepsilon}{c}\frac{\partial\mathbf{E}}{\partial t},$$

where the magnetic field determined by the potentials (2.10) is used for the left-hand side of this equation. It can be confirmed that the equation obtained for the phase difference in this case is the same as the one obtained by our approach, which is taken from Refs. 12 and 13 and is based on the condition of continuity of the current.

Using the function ψ (2.16), we can write the imaginary part of the Fourier component of the kernel $Q''(k_z,\omega)$ in the form

$$Q''(k_z,\omega)=\left\{\frac{1}{a^2+1}\left[a\sqrt{a^2\omega^2+\omega^2-c^2k_z^2}+\frac{c^2k_z^2}{\sqrt{a^2\omega^2+\omega^2-a^2c^2k_z^2}}\right]-|\omega|\right\}\frac{\omega}{2c^2}\times\eta(\omega^2-c^2k_z^2), \tag{2.25}$$

where

$$a \equiv 2\lambda_e|\omega|/c. \tag{2.26}$$

The expression for the real part of the Fourier component of the kernel $Q'(k_z,\omega)$ is more cumbersome (see Appendix B), but it has comparatively simple asymptotic representations in important limiting cases. In particular, for $|\omega| \ll c|k_z|$ and $a \ll 1$ we obtain the following from Eqs. (B1) and (B6):

$$Q'(k_z,\omega) = \frac{k_z^2}{\pi} Q(2\lambda_e|k_z|) - \frac{\omega^2}{\pi c^2} \ln \frac{1}{d|k_z|}, \tag{2.27}$$

where the function $Q(x)$ has the form[14]

$$Q(x) = \frac{2}{\sqrt{x^2-1}} \arctan \frac{\sqrt{x^2-1}}{1+x}. \tag{2.28}$$

In the low-frequency limit, when the second term in Eq. (2.27) can be disregarded, the Fourier component of the kernel $Q(k_z,\omega)$ (2.24), (2.27) goes over to the well-known result of Ref. 14. In the general case, in contrast with Ref. 14, $Q(k_z,\omega)$ depends on the frequency. Frequency dispersion sets in as a result of the sequential description of the electromagnetic potentials in vacuum and can be exploited, in particular, to investigate the influence of the vacuum field on the dispersion properties of a Josephson junction. Frequency dispersion is especially conspicuous in the limit $|\omega| \gg c|k_z|$.

In this limit, but for $\alpha \ll 1$, we obtain the following from Eqs. (B1) and (B7):

$$Q'(k_z,\omega) = \frac{k_z^2}{\pi} \ln \frac{c}{\lambda_e|\omega|} - \frac{\omega^2}{\pi c^2} \ln \frac{c}{d|\omega|}. \tag{2.29}$$

We note that Eqs. (2.27) and (2.29) are well matched for $|\omega| \sim c|k_z|$. This means that for $\alpha \ll 1$ Eqs. (2.27) and (2.29) are sufficient for describing the whole range of the parameter $|\omega/ck_z|$.

In closing this section, we emphasize that Eq. (2.22) takes into account Ohmic losses of the current flowing across the Josephson junction in accordance with Eq. (2.19). This property is consistent with the model of Refs. 16 and 17 and its elaboration in Ref. 18. The equation does not take into account Ohmic losses in superconductors due to the presence of normal electrons, whose inclusion in Ref. 19 makes it possible to broaden the nonlocal description of dissipation in Josephson junctions with a high critical current density. Our disregard for he contribution of normal electrons in superconductors is admissible, first, at sufficiently low temperatures, when the fraction of normal electrons is small, and, second, when the states in question are found to have sufficiently long wavelengths, as is true in the problems of wave emission from a Josephson junction discussed below.

3. SWIHART WAVES

We use the temporally and spatially nonlocal equations (2.22)–(2.24) to analyze waves traveling along the junction and having the form

$$\varphi(z,t) = \frac{1}{2}\varphi_0 \exp(ik_z z - i\omega t) + \text{c.c}, \quad k_z > 0, \tag{3.1}$$

where $|\varphi_0| \ll 1$. Replacing $\sin\varphi$ by the argument φ, we then obtain a dispersion relation describing the relationship between the frequency ω and the component of the wave vector k_z:

$$\omega^2 + i\beta\omega - \omega_j^2 = 4\frac{d}{\varepsilon D} c^2 [Q'(k_z,\omega) + iQ''(k_z,\omega)]. \tag{3.2}$$

We now examine the consequences of Eq. (3.2) in the most interesting case, when

$$|\omega| \ll c/2\lambda_e = cD/2\lambda^2. \tag{3.3}$$

If the typical values of the London depth are $\lambda \sim 10^{-5}$ cm and the thickness of the superconducting film $D \sim 3 \times 10^{-7}$ cm, inequality (3.3) is satisfied for $|\omega| \ll 5 \times 10^{13}\,\text{s}^{-1}$. Besides this condition, as usual, the following inequality is also assumed to hold:

$$|\omega| \ll \Delta/\hbar \tag{3.4}$$

(Δ is the width of the superconducting gap), which gives $|\omega| \ll 1.4 \times 10^{13}\,\text{s}^{-1}$ for $\Delta \sim 100$ K.

We first discuss the consequences of Eq. (3.2) in the limit of Swihart wave phase velocities much lower than the speed of light,

$$|\omega/k_z| \ll c. \tag{3.5}$$

In this limit we have $Q''(k_z,\omega) = 0$ [Eq. (2.25)], and the real part of the Fourier component of the kernel $Q'(k_z,\omega)$ is described by Eqs. (2.27) and (2.28). Substituting Eq. (2.27) into (3.2) and assuming that $\omega \equiv \omega' - i\gamma \simeq \omega'$, which is valid for small β, we obtain a relation between the real part of the frequency ω' and the component of the wave vector k_z:

$$\omega'^2 = \omega_j^2 \frac{1 + \frac{4}{\pi} k_z^2 l \lambda_e Q(2\lambda_e k_z)}{\Lambda(k_z)}, \tag{3.6}$$

where the function $\Lambda(k_z)$ exhibits a logarithmically weak dependence on k_z:

$$\Lambda(k_z) \equiv 1 + \frac{4}{\pi}\frac{d}{\varepsilon D} \ln\left(\frac{1}{dk_z}\right). \tag{3.7}$$

The dispersion law (3.6) differs from the standard version[14,20] by the function $\Lambda(k_z)$ in the denominator; its presence can be interpreted as a consequence of the influence of the surface electromagnetic wave (2.17) on the Swihart wave. For typical Josephson junctions in a thin film $2d$ and D are of the same order and $\varepsilon \sim 2$, so that $4d/\pi\varepsilon D \sim 0.3$. On the other hand, the characteristic scales of variation of the phase difference along the y axis are of the order of λ_e, i.e., $\ln(1/dk_z) \sim \ln(\lambda_e/d) \sim 7$. We ultimately obtain $\Lambda \sim 3$. The surface electromagnetic wave also influences the damping of the Swihart wave due to Ohmic energy losses in the tunnel junction. Invoking the dispersion law (3.6), from (3.2) we obtain the following inequality for the corresponding damping constant γ:

$$\gamma = \beta/2\Lambda(k_z) < \beta/2. \tag{3.8}$$

We now analyze Eq. (3.2) in the opposite limit from (3.5), when the phase velocity of the Swihart wave is much greater than that of light:

$$|\omega/k_z| \gg c. \tag{3.9}$$

Bearing relations (2.25) and (2.29) in mind and assuming that $\omega' \gg \gamma$, we then obtain an equation for the relation between ω' and k_z:

$$\omega'^2 = \omega_j^2 + \frac{4}{\pi}\frac{d}{\varepsilon D}c^2\left\{k_z^2 \ln\frac{c}{\lambda_e\omega'} - \frac{\omega'^2}{c^2}\ln\frac{c}{d\omega'}\right\}. \tag{3.10}$$

Inasmuch as the frequency ω' is close to the Josephson frequency in the range of relativistic phase velocities, we obtain the following from Eq. (3.10) to logarithmic accuracy:

$$\omega'^2 = \omega_j^2 \frac{1 + \frac{4}{\pi}k_z^2 l\lambda_e \ln\frac{c}{\lambda_e\omega_j}}{\Lambda(\omega_j/c)},. \tag{3.11}$$

We note that for $\omega' \sim ck_z$, by virtue of condition (3.3), the dispersion law (3.11) goes over to the law (3.6), which holds for low Swihart wave phase velocities.

We calculate the small damping constant by means of Eq. (2.25), which under the conditions of inequality (3.3) and

$$\omega^2 - c^2k_z^2 \gg a^2\omega^2 \tag{3.12}$$

has the form

$$Q''(k_z,\omega) = \frac{1}{2}\left(k_z^2 - \frac{\omega^2}{c^2}\right)\operatorname{sgn}\omega. \tag{3.13}$$

Then for $\gamma \ll \omega'$ the following can be obtained from Eq. (3.2):

$$\gamma = \frac{\frac{\beta}{2} + \frac{d}{\varepsilon D\omega'}(\omega'^2 - c^2k_z^2)}{\Lambda(\omega'/c)}, \tag{3.14}$$

where the frequency ω' is described by Eq. (3.11).

If the opposite condition holds instead of (3.12), i.e., if

$$a^2\omega^2 \gg \omega^2 - c^2k_z^2 \geqslant 0, \tag{3.15}$$

then the imaginary part $Q''(k_z,\omega)$ (2.25) is half the value given by Eq. (3.13). This means that for phase velocities close to the speed of light the term without β in (3.14) must be reduced by one half.

According to Eq. (3.14), for Swihart waves with phase velocities above the speed of light the damping constant contains two additive contributions. The first contribution to γ is proportional to β, exists for $\omega' < ck_z$, and is attributable to Ohmic energy losses in the tunnel junction. The second contribution to γ contains the factor $\omega'^2 - c^2k_z^2$, exists only for high phase velocities, and describing the damping of Swihart waves due to energy losses in the emission of electromagnetic waves into vacuum. In the long-wavelength range, where $k_z \ll \omega_j/c$, the second contribution is the main term if the conductivity of the tunnel junction is sufficiently low,

$$\sigma < \frac{d}{2\pi D}\omega_j, \tag{3.16}$$

which is equivalent to the inequality $\beta < (2d/\varepsilon D)\omega_j$. If we set $2d \sim D$ and $\varepsilon \sim 10$, the condition $\beta < 0.1\omega_j$ must be satisfied. We note that Eq. (3.14) has been obtained in the limit $\gamma \ll \omega'$. Comparing relations (3.11) and (3.14), for $\omega' \gg ck_z$ and $\beta \ll 2\omega' d/\varepsilon D$ we infer that $\gamma \ll \omega'$ holds if

$$d \ll \varepsilon D\Lambda(\omega'/c) \simeq \varepsilon D\Lambda(\omega_j/c). \tag{3.17}$$

Consequently, this behavior of the spectrum of Swihart waves occurs in films having thicknesses in the interval $\lambda \gg D \gg d/\varepsilon\Lambda(\omega_j/c)$.

4. EMISSION FROM LINEAR WAVES

Here we discuss the characteristics of radiation from a long Josephson junction when a linear phase-difference wave of the form (3.1) propagates along it. We assume that the phase velocity of the wave is much greater than the speed of light, $\omega/k_z > c$. According to Eqs. (2.10)–(2.16), the only Fourier components of the potential that are emitted are those for which

$$k_x^2 < \omega^2/c^2 - k_z^2, \tag{4.1}$$

and the function ψ (2.16) is purely imaginary. The contribution of the scalar potential (2.14) can now be disregarded, because it does not produce any radiative energy losses and is negligible far from the film. Invoking Eqs. (2.10)–(2.16), we can then find the components of the vector potential in vacuum:

$$A_x(\mathbf{r},t) = -\frac{\phi_0\varphi_0}{2\pi^2}\int_0^\kappa dk_x \cos(xk_x)\{[Z_2 + k_x^2 \times (1 - a^2 - Z_1)]\cos\alpha - 2\lambda_e\sqrt{\kappa^2 - k_x^2} \times [Z_2 + k_x^2(1 - a^2)]\sin\alpha\}(Z_1Z_2)^{-1}, \tag{4.2}$$

$$A_y(\mathbf{r},t) = -\frac{\phi_0\varphi_0}{2\pi^2}\int_0^\kappa dk_x k_x \sin(xk_x) \times \left\{a\frac{\omega}{c}\cos\alpha + \sqrt{\kappa^2 - k_x^2}\sin\alpha\right\}Z_2^{-1}, \tag{4.3}$$

$$A_z(\mathbf{r},t) = -\frac{\phi_0\varphi_0}{\pi^2}\lambda_e k_z\int_0^\kappa dk_x k_x \times \sin(xk_x)\left\{2\lambda_e\left[Z_2 + \frac{\omega^2}{c^2}(1 - a^2)\right] \times \sin\alpha - (1 - a^2)\sqrt{\kappa^2 - k_x^2}\cos\alpha\right\}(Z_1Z_2)^{-1}, \tag{4.4}$$

where the following notation has been introduced:

$$\kappa^2 \equiv \omega^2/c^2 - k_x^2,$$
$$\alpha \equiv \sqrt{\kappa^2 - k_x^2}\,y + k_z z - \omega t, \tag{4.5}$$
$$Z_1 \equiv 1 + 4\lambda_e^2(\kappa^2 - k_x^2), \quad Z_2 \equiv \kappa^2 - k_x^2 + \alpha^2\omega^2/c^2. \tag{4.6}$$

Equations (4.2)–(4.4) can be used to find the contribution to the electromagnetic field components **E** and **H** from the radiation of waves into vacuum:

$$\mathbf{E}=-\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{H}=\operatorname{curl}\mathbf{A}. \tag{4.7}$$

In the equation for **E** (4.7), as in the derivation of Eqs. (4.2)–(4.4), the term representing the contribution of the scalar potential, $-\operatorname{grad} V$, has been omitted. Using the radiation-descriptive parts of the fields **E** and **H**, we can find the electromagnetic radiation flux density averaged over the period of the field:

$$\mathbf{S}=\frac{\omega}{2\pi}\int_0^{2\pi/\omega} dt\,\frac{1}{4\pi}\left[\operatorname{curl}\mathbf{A}\frac{\partial \mathbf{A}}{\partial t}\right]. \tag{4.8}$$

Substituting Eqs. (4.2)–(4.4) into (4.8) and evaluating the integrals with respect to the variable k_x by a procedure similar to that described in Appendix C, we find

$$S_x=S_0R\cos\xi, \tag{4.9}$$

$$S_y=S_0R\sin\xi, \tag{4.10}$$

$$S_z=\frac{S_0R}{\sqrt{v^2-1}}, \tag{4.11}$$

where $\xi\equiv\arctan(y/x)$ varies in the interval from 0 to 2π. In Eq. (4.9) S_0 characterizes the order of magnitude of the flux at the distance $r_\perp=\sqrt{x^2+y^2}$ from the Josephson junction:

$$S_0\equiv\left(\frac{\phi_0\varphi_0\omega}{8\pi^2c}\right)^2\frac{\omega}{r_\perp}, \tag{4.12}$$

the function R depends on the angle ξ and the ratio of the Swihart wave phase velocity to the speed of light, $v=\omega/ck_z$:

$$R\equiv\left[1+\frac{a^2-1}{v^2+a^2(v^2-1)\sin^2\xi}\right]\left[1+\frac{a^2v^2}{(v^2-1)\sin^2\xi}\right]^{-1}. \tag{4.13}$$

From Eqs. (4.9)–(4.11) we obtain an equation for the ratio of the component of the flux density along the Josephson junction, S_z, to the flux density in the plane normal to the axis of the tunnel junction, $S_\perp\equiv\sqrt{S_x^2+S_y^2}=S_0R$:

$$\frac{S_z}{S_\perp}=\frac{1}{\sqrt{v^2-1}}. \tag{4.14}$$

According to Eq. (4.14), $S_z\gg S_\perp$ for $1\gg v^2-1>0$, and electromagnetic energy is emitted in the direction of propagation of the Swihart wave. On the other hand, for $v\gg1$, when the phase velocity is much greater than the speed of light, $S_\perp\gg S_z$, and the radiation is localized mainly in a plane orthogonal to the axis of the tunnel junction. From Eqs. (4.9)–(4.11) we can obtain expressions for the total flux density of the radiation at an angle ξ relative to the plane of the film:

$$S=\sqrt{S_z^2+S_\perp^2}=S_0R\frac{v}{\sqrt{v^2-1}}. \tag{4.15}$$

Bearing in mind the smallness of the parameter a, from Eqs. (4.13) and (4.15) we have

$$\frac{S}{S_0}=\frac{v^2-1+a^2}{v\sqrt{v^2-1}}\left[1+\frac{a^2v^2}{(v^2-1)\sin^2\xi}\right]^{-1}. \tag{4.16}$$

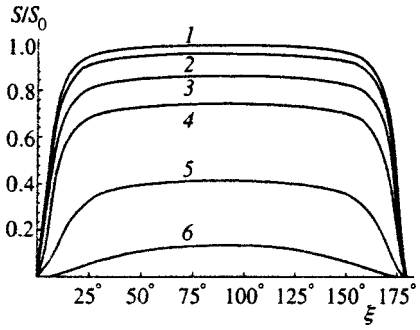

FIG. 1. Radiation flux density in a half-space over a film as a function of the angle of observation ξ for various values of the parameter: (*1*) $(\omega/ck_z)^2=100$; (*2*) 40; (*3*) 2; (*4*) 1.5; (*5*) 1.1; (*6*) 1.01.

It is evident from this result that for $1\geqslant|\sin\xi|>av/\sqrt{v^2-1}$ the energy flux density S is essentially independent of the angle ξ. But for $|\sin\xi|\ll av/\sqrt{v^2-1}$

$$\frac{S}{S_0}\simeq\frac{\sqrt{v^2-1}}{v}(v^2-1+a^2)\frac{\sin^2\xi}{a^2v^2} \tag{4.17}$$

varies as a function of $\sin^2\xi$.

These functional relations are illustrated in Fig. 1, which shows the dependence of the function S/S_0 on the angle ξ. The curves have been plotted on the assumption that $a=2\lambda_e|\omega|/c=0.1$. The graphs of S/S_0 are given for six values of the parameter v^2. It is evident in Fig. 1 how the angular range in which the radiation flux density is constant broadens as the phase velocity increases. The dependence of S/S_0 on the ratio of ω/k_z to the speed of light c is shown in Fig. 2. The curves in this figure correspond to five different angles of observation ξ. According to Fig. 2, the smaller $|\sin\xi|$, the smaller is the radiation flux density. The flux density S/S_0 tends monotonically to zero as ω/ck_z approaches unity.

Equations (4.9)–(4.11) can also be used to find the damping constant of waves traveling along the Josephson junction due to radiative energy losses. In fact, the energy losses per unit length of the tunnel junction are given by the integral

$$r_\perp\int_0^{2\pi}d\xi\,\mathbf{n}\cdot\mathbf{S}=r_\perp S_0\int_0^{2\pi}d\xi R=-\frac{4\pi c^2}{\omega^2}r_\perp S_0Q''(k_z,\omega), \tag{4.18}$$

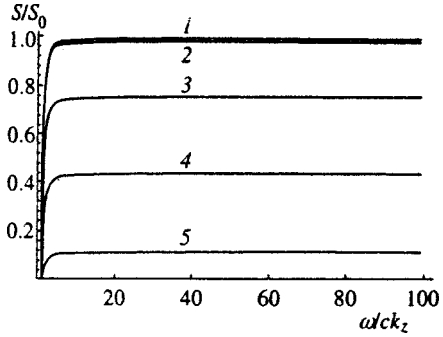

FIG. 2. Dependence of the function S/S_0 on ω/ck_z for various angles of observation: (*1*) 90°; (*2*) 45°; (*3*) 10°; (*4*) 5°; (*6*) 2°.

where **n** is the unit vector normal to the cylindrical surface of radius $r_\perp=\sqrt{x^2+y^2}$, and the imaginary part $Q''(k_z,\omega)$ is described by Eq. (2.25).

On the other hand, the radiative energy losses can be written in the form $2\gamma_r W$, where γ_r is the radiative damping constant, and W is the period-average energy per unit length of the Josephson junction. To calculate W for waves with a phase velocity greater than the speed of light and under the condition

$$\frac{d}{\varepsilon D}\ln\frac{c}{d\omega_j}\ll 1, \tag{4.19}$$

when the difference of the wave frequency ω from the Josephson frequency ω_j can be ignored, it is sufficient to take into account the energy of the electric field in the junction and the energy of the Josephson current. We then have

$$W=\frac{\omega}{2\pi}\int_0^{2\pi/\omega}dt\,\frac{\hbar j_c}{2|e|}D\left[\frac{1}{2\omega_j^2}\left(\frac{\partial}{\partial t}\varphi\right)^2+1-\cos\varphi\right]\simeq\frac{\varepsilon D}{d}S_0\frac{\pi r_\perp}{\omega}. \tag{4.20}$$

Setting $2\gamma_r W$ equal to the radiative energy losses (4.18), we obtain

$$\gamma_r=-\frac{2d}{\varepsilon D}\frac{c^2}{\omega}Q''(k_z,\omega), \tag{4.21}$$

which is consistent with Eqs. (3.13) and (3.14).

5. EMISSION OF LARGE-SCALE NONLINEAR STATES

In this section we discuss the emission of waves into vacuum from nonlinear states when the amplitude of the phase difference of the wave functions is not small in comparison with unity. We assume that the variation of the phase difference has a long characteristic time:

$$T\equiv\left|\frac{\partial}{\partial t}\ln\varphi(z,t)\right|^{-1}\gg\frac{2\lambda_e}{c}. \tag{5.1}$$

We assume that the space scale of variation of the phase difference L is much greater than the distance traversed by the electromagnetic wave in vacuum during the characteristic time of variation of the phase difference:

$$L\equiv\left|\frac{\partial}{\partial z}\ln\varphi(z,t)\right|\gg cT. \tag{5.2}$$

Inequalities (5.1) and (5.2) are compatible if the inhomogeneity scale L is much greater than twice the effective penetration depth of the field into the film, $L\gg 2\lambda_e$. We refer to states with space scales L that satisfy inequality (5.2) as large-scale states. In addition to conditions (5.1) and (5.2), we assume that

$$\frac{d}{\varepsilon D}\ln\frac{cT}{d}\ll 1. \tag{5.3}$$

Inequality (5.3) permits the influence of a surface electromagnetic wave on a large-scale nonlinear state to be ignored. Under the conditions defined by inequalities (5.1)–(5.3), taking Eqs. (2.24), (2.25), and (2.29) into account, we can write Eq. (2.22) in the form

$$\sin\varphi(z,t)+\frac{\beta}{\omega_j^2}\frac{\partial}{\partial t}\varphi(z,t)+\frac{1}{\omega_j^2}\frac{\partial^2}{\partial t^2}\varphi(z,t)=-\frac{2}{\pi}\lambda_e l\frac{\partial^2}{\partial z^2}\int_{-\infty}^{\infty}dt'\,\mathrm{sgn}\,(t'-t)\times\left[C+\ln\left(\frac{c}{\lambda_e}|t'-t|\right)\right]\frac{\partial}{\partial t'}\varphi(z,t')-\frac{2}{\pi}\lambda_e l\frac{\partial}{\partial t}\int_{-\infty}^{\infty}\frac{dt'}{t-t'}\frac{\partial}{\partial t'}\varphi(z,t'). \tag{5.4}$$

where $C=0.577$ is the Euler constant. Unlike Eq. (2.22), Eq. (5.4) does not contain spatially nonlocal coupling. When the characteristic time of variation of the phase difference is determined by the reciprocal Josephson frequency, the argument of the logarithm on the right-hand side of (5.4) can be approximately replaced by $c/\lambda_e\omega_j$. Equation (5.4) now assumes the form

$$\omega_j^2\sin\varphi(z,t)+\beta\frac{\partial}{\partial t}\varphi(z,t)+\frac{\partial^2}{\partial t^2}\varphi(z,t)=\frac{4}{\pi}\frac{d}{\varepsilon D}c^2\ln\left(\frac{c}{\lambda_e\omega_j}\right)\frac{\partial^2}{\partial z^2}\varphi(z,t)+\frac{2}{\pi}\frac{d}{\varepsilon D}\frac{\partial}{\partial t}\int_{-\infty}^{\infty}\frac{dt'}{t'-t}\frac{\partial}{\partial t'}\varphi(z,t') \tag{5.5}$$

We note that comparatively small terms occur on the right-hand sides of Eqs. (5.4) and (5.5) by virtue of the stated assumptions (5.2) and (5.3). However, these terms govern both the radiative energy losses and the slow spatial variation of the phase difference. In particular, these terms characterize the dispersion and radiation damping of Swihart waves (see Sec. III). We now use Eq. (5.4) [or (5.5)] to analyze the emission of waves by large-scale distributions of the phase difference. Taking inequalities (5.2) and (5.3) into account, we disregard the influence of the small terms on the right-hand side of (5.4) in the first approximation.

We also assume that the dissipation associated with Ohmic losses is small: $\beta\ll\omega_j$. Under these conditions Eq. (5.4) reduces to the mathematical pendulum equation

$$\omega_j^2\sin\varphi(z,t)+\frac{\partial^2}{\partial t^2}\varphi(z,t)=0, \tag{5.6}$$

whose solutions are well known (see, e.g., Refs. 3 and 21). The radiative energy losses by nonlinear states described by the solutions of Eq. (5.6) have been investigated in a brief communication.[21] We emphasize that even though the derivative of the phase difference with respect to the coordinate for such states is equal to zero, the electromagnetic field on the surface of the superconducting film is not equal to zero. The presence of this field induces a radiation flux that transports energy away from the junction into vacuum (see Appendix D). We now discuss the specific characteristics of the emission of waves by large-scale nonlinear states in greater

detail. Since the distribution of the phase difference of such states is independent of the coordinate in the first approximation, the Fourier component $\varphi(k_z,\omega)$ is proportional to the delta function $\delta(k_z)$. Invoking Eqs. (2.10)–(2.16), disregarding the contribution of the scalar potential, and retaining only the emitted Fourier components of the potentials, we find

$$A_x(\mathbf{r},t)=-\frac{\phi_0}{4\pi^3}\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}dt'\exp[i\omega(t'-t)]\varphi(t')\times\int_0^{|\omega|/c}dk_x\cos(xk_x)\exp[-y\psi(\omega,k_x)]\times\frac{\psi(\omega,k_x)}{\psi(\omega,k_x)-a\omega/c},\tag{5.7}$$

$$A_y(\mathbf{r},t)=-\frac{\phi_0}{4\pi^3}\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}dt'\times\exp[i\omega(t'-t)]\varphi(t')\int_0^{|\omega|/c}dk_xk_x\sin(xk_x)\times\exp[-y\psi(\omega,k_x)]\frac{1}{\psi(\omega,k_x)-a\omega/c},\tag{5.8}$$

where $\varphi(t)$ is the solution of Eq. (5.6),

$$\varphi(\omega,k_x)=-i\sqrt{\omega^2/c^2-k_x^2}\ \operatorname{sgn}\omega.\tag{5.9}$$

In this approximation we have $A_z(\mathbf{r},t)=0$.

We first discuss the emission of waves for the nonlinear state described by the solutions of Eq. (5.6) corresponding to finite-amplitude oscillations of the phase difference, when

$$\varphi_v(t)=2\arcsin[k\,\mathrm{sn}(\omega_jt,k)],\tag{5.10}$$

where k is the modulus of the elliptic sine sn. We expand the derivative of the function φ_v into a Fourier series:

$$\frac{d}{dt}\varphi_v(t)=8\Omega_v\sum_{n=0}^{\infty}\frac{q^{n+1/2}}{1+q^{2n+1}}\cos[(2n+1)\Omega_vt],\tag{5.11}$$

where the frequency Ω_v and the parameter q depend on the modulus k:

$$\Omega_v=\frac{\pi\omega_j}{2K(k)},\tag{5.12}$$

$$q=\exp\left[-\frac{\pi}{K(k)}K(\sqrt{1-k^2})\right],\tag{5.13}$$

and $K(k)$ is a complete elliptic integral of the first kind. Now, using the expressions for the components of the vector potential (5.7) and (5.8) and the expansion (5.11), we can find the radiation flux density averaged over the period $2\pi/\Omega_v$ [cf. (4.8)]:

$$\mathbf{S}=(\mathbf{e}_x\cos\xi+\mathbf{e}_y\sin\xi)\sum_{n=0}^{\infty}\frac{S_{v,n}\sin^2\xi}{\sin^2\xi+[2\lambda_e(2n+1)\Omega_v/c]^2},\tag{5.14}$$

where we have introduced the notation

$$S_{v,n}=\frac{\phi_0^2\Omega_v^3}{\pi^4c^2r_\perp}\frac{q^{2n+1}}{(1+q^{2n+1})^2}(2n+1).\tag{5.15}$$

According to Eqs. (5.11) and (5.14), radiation is emitted in odd harmonics of the fundamental Ω_v (5.12), which is smaller than the Josephson frequency. For $k\ll1$ the frequency Ω_v is close to ω_j, and the radiative energy losses are governed mainly by emission from the fundamental $n=0$. As k increases, the radiation spectrum becomes more abundant in higher harmonics, and the interval between the harmonics decreases. When k is close to unity, Ω_v tends logarithmically to zero:

$$\Omega_v\simeq\frac{\pi}{2}\omega_j\left(\ln\frac{4}{\sqrt{1-k^2}}\right)^{-1}.\tag{5.16}$$

with the emission of a broad spectrum of closely spaced frequencies $(2n+1)\Omega_v$. We note, in accordance with the conditions (5.1) and (5.2) underlying the validity of the theory, that the emitted frequencies must lie in the interval

$$2\pi c/L\ll(2n+1)\Omega_v\ll\pi c/\lambda_e.\tag{5.17}$$

Equation (5.14) can be used to express the directionality of the emission of harmonics. According to Eq. (5.14), the radiation flux from large-scale nonlinear states is localized in a plane orthogonal to the axis of the Josephson junction and is directed along the radius vector $\mathbf{r}_\perp$. Since the parameter $2\lambda_e(2n+1)\Omega_v/c$, by virtue of the right inequality (5.17), is much smaller than unity over a wide range of angles, when

$$|\sin\xi|>2(2n+1)\lambda_e\Omega_v/c,\tag{5.18}$$

the absolute value of the flux density at the frequency $(2n+1)\Omega_v$ does not depend on the angle of observation and is equal to $S_{v,n}$. Outside the interval (5.18), as when the angles of observation are tight against the plane of the film, the intensity of emission of the harmonics decays as $|\sin\xi|$ decreases, in proportion to $\sin^2\xi$. Integrating the flux $\mathbf{S}$ (5.14) over the surface of a cylinder of radius $r_\perp$ and taking into account the smallness of the parameter $2(2n+1)\lambda_e\Omega_v/c$, we obtain Eq. (11) in Ref. 21 for the energy flux from unit length of the tunnel junction.

Next we consider the emission of radiation from the nonlinear state described by the solution of the rotating pendulum equation (5.6), when

$$\varphi_r(t)=2\,\mathrm{am}\left(\frac{\omega_j}{k}t,k\right)=2\Omega_rt+4\sum_{n=1}^{\infty}\frac{q^n}{n(1+q^{2n})}\sin[2n\Omega_rt],\tag{5.19}$$

where am denotes the Jacobi amplitude, and

$$\Omega_r=\frac{\pi}{2kK(k)}\omega_j.\tag{5.20}$$

In this case the radiation flux density averaged over the period $2\pi/\Omega_r$ has the form

$$\mathbf{S}=(\mathbf{e}_x\cos\xi+\mathbf{e}_y\sin\xi)\sum_{n=1}^{\infty}\frac{S_{r,n}\sin^2\xi}{\sin^2\xi+(4n\lambda_e\Omega_r/c)^2},\tag{5.21}$$

$$S_{r,n}=\frac{\phi_0^2\Omega_r^3}{\pi^4c^2r_\perp}\frac{2nq^{2n}}{(1+q^{2n})^2}. \tag{5.22}$$

Equations (5.21) and (5.22) describe the radiation flux at even frequencies $2n\Omega_r$. If k is close to unity, the frequency Ω_r (5.20) is much lower than the Josephson frequency, and the radiation spectrum contains many even harmonics of the fundamental. The frequency Ω_r becomes lower as k increases. If $k\ll 1$, we have $\Omega_r\simeq\omega_j/k\gg\omega_j$. The intensity of the radiation at higher harmonics is significantly reduced since $q\ll 1$. The radiation pattern of the even harmonics is the same as for odd harmonics. The radiation flux is localized in a plane orthogonal to the axis of the tunnel junction. If $|\sin\xi|>4n\lambda_e\Omega_r/c$, the flux density at the frequency $2n\Omega_r$ does not depend on the angle of observation and is equal to $S_{r,n}$ (5.22). Another solution of Eq. (5.6) is the 2π kink

$$\varphi(t)=4\arctan(e^{\omega_j t})-\pi. \tag{5.23}$$

The radiation pattern from the 2π kink has the same features as for nonlinear periodic solutions. Now a continuous spectrum of frequencies lower than, or of the same order as the Josephson frequency is emitted. At high frequencies the spectral density of the radiation is exponentially small.[21]

6. CONCLUSION

To summarize the foregoing discussion, the space–time nonlocal electrodynamics describing transient states in a Josephson contact in a superconducting thin film has enabled us to obtain comparatively simple and transparent equations that provide a picture of electromagnetic radiation penetrating from the surface of the film into vacuum. This possibility has been afforded both by the formulation of a nonlinear integrodifferential equation describing the distribution of the phase difference in the junction and by the systematic description of the electromagnetic field in vacuum.

However, the above treatment ignores dissipation associated with thermal fluctuations. We are grateful to the reviewer who called our attention to Ref. 22, in which the authors have investigated the influence of voltage fluctuations at the contact on the width of the emission line from a Josephson point contact to which a constant voltage V_0 is applied. An equation for the width of the emission line at the frequency $2eV_0/\hbar$ can be deduced from Ref. 22:

$$\Gamma=\left(\frac{4\pi c}{\phi_0}\right)^2\frac{2\pi d}{\beta\varepsilon S_{\rm con}}T_{\rm con}, \tag{6.1}$$

where $T_{\rm con}$ is the temperature, and $S_{\rm con}$ is the area of the point contact. Equation (6.1) is valid when

$$\Gamma\ll\beta, \tag{6.2}$$

and the external circuit has a high resistance.

It is useful to ascertain the conditions under which the influence of fluctuations can be disregarded. According to Eq. (2.22), the contribution of radiation to the linewidth is determined additively by the term β; hence, our analysis is appropriate when inequality (6.2) is satisfied.

It should be noted here, in accordance with Ref. 1, p. 114, and Ref. 2, p. 72, that the most significant conditions in practical applications of point contacts correspond to the satisfaction of inequality (6.2). Indeed, assuming, for example, that we have $T=4.2$ K, $\beta\simeq 5\times 10^{10}\,\mathrm{s}^{-1}$, and $2d/\varepsilon\simeq 3\times 10^{-8}$ cm, we see that condition (6.2) is satisfied for point contacts having an area greater than $\sim 10\,\mu\mathrm{m}^2$. In application to our case of a contact in a film, $S_{\rm con}=DL_f$, where L_f is the width of the film, condition (6.2) is satisfied when the width of the film is not too small. For example, adopting the same tunnel junction parameters as above, along with $D\sim 3\times 10^{-7}$ cm, we find that inequality (6.2) is satisfied for $L_f>3$ mm. Consequently, it is clearly admissible to neglect the influence of fluctuation dissipation in comparison with ordinary Ohmic dissipation.

Finally, it must be emphasized that our emission of radiation associated with Swihart waves is possible at frequencies comparable with the Josephson frequency ω_j. When condition (6.2) is satisfied in this case, emission can be efficient under the condition

$$\omega_j\gg\beta. \tag{6.3}$$

Condition (6.3) is satisfied for the typical parameters $\omega_j\sim 10^{12}\,\mathrm{s}^{-1}$ and $\beta\sim 5\times 10^{10}\,\mathrm{s}^{-1}$ (see, e.g., Ref. 23).

However, for a film of sufficiently small width, such that condition (6.2) is violated, fluctuations can in fact dictate the emission linewidth. In this limit, according to Ref. 22, the emission linewidth is given by the expression

$$\Gamma=\frac{8|e|}{\hbar}\sqrt{\frac{\pi d}{\varepsilon S_{\rm con}}T_{\rm con}}. \tag{6.4}$$

It is clear that the emission is adequately efficient if

$$\frac{8|e|}{\hbar}\sqrt{\frac{\pi d}{\varepsilon S_{\rm con}}T_{\rm con}}\ll\omega_j.$$

This inequality, like inequality (6.2), imposes restrictions on the width of the film. For the Josephson junction parameters adopted above, it is satisfied if $L_f\gg 10\,\mu$m.

The latter result indicates that our above discussion of the conditions for inequality (6.2) to hold is of methodological significance only, which, of course, is important in understanding the validity of the model used in our article.

In summary, our investigation has established laws governing the emission of electromagnetic radiation from the surface of a film containing a Josephson junction.

This work has been performed as part of Project 96-02-17303 of the Russian Fund for Fundamental Research with support from the Scientific Council on High-Temperature Superconductors (Project ''AD'' 95008) and state support for leading scientific schools (Project 96-15-96750).

APPENDIX A:

To construct a solution of Eqs. (2.1)–(2.4), we go over to Fourier transforms in the variables $\mathbf{r}_\parallel$ and t:

$$F(y,\mathbf{k},\omega)\equiv\int_{-\infty}^{\infty}dt\int d\mathbf{r}_\parallel\exp(i\omega t-i\mathbf{k}\cdot\mathbf{r}_\parallel)F(y,\mathbf{r}_\parallel,t), \tag{A1}$$

where $\mathbf{k}=(k_x,0,k_z)$. We can then obtain a system of equations for the Fourier components of the potentials:

$$\frac{d}{dy}A_y(y,\mathbf{k},\omega)+i\mathbf{k}\cdot\mathbf{A}_\parallel(y,\mathbf{k},\omega)=0, \tag{A2}$$

$$\frac{d^2}{dy^2}\mathbf{A}(y,\mathbf{k},\omega)-\left(k^2-\frac{\omega^2}{c^2}\right)\mathbf{A}(y,\mathbf{k},\omega) +i\frac{\omega}{c}\left(i\mathbf{k}+\mathbf{e}_y\frac{d}{dy}\right)V(y,\mathbf{k},\omega)=0,\quad y>0, \tag{A3}$$

$$\frac{d^2}{dy^2}V(y,\mathbf{k},\omega)-k^2V(y,\mathbf{k},\omega)=0. \tag{A4}$$

Here the boundary conditions (2.8) and (2.9) for the components of the vector potential have the form

$$\frac{d}{dy}A_x(y,\mathbf{k},\omega)\big|_{y=+0}=\frac{1}{2\lambda_e}\left\{\frac{\phi_0}{2\pi}[\varphi(k_z,\omega)+ik_x\Phi(\mathbf{k},\omega)] +A_x(y=0,\mathbf{k},\omega)\right\}, \tag{A5}$$

$$\frac{d}{dy}A_z(y,\mathbf{k},\omega)\big|_{y=+0}=\frac{1}{2\lambda_e}\left\{\frac{\phi_0}{2\pi}ik_z\Phi(\mathbf{k},\omega) +A_z(y=0,\mathbf{k},\omega)\right\} \tag{A6}$$

$$A_y(y=0,\mathbf{k},\omega)=0, \tag{A7}$$

where $\mathbf{e}_x$ and $\mathbf{e}_y$ are unit vectors along the x and y axes, and the Fourier components of the phase and the phase difference are described by the relations

$$\Phi(\mathbf{k},\omega)=\int_{-\infty}^{0}dx\Phi_1(x,k_z,\omega)\exp(-ik_xx) +\int_{-\infty}^{\infty}dx\Phi_2(x,k_z,\omega)\exp(-ik_xx), \tag{A8}$$

$$\varphi(k_z,\omega)=\Phi_1(0,k_z,\omega)-\Phi_2(0,k_z,\omega). \tag{A9}$$

It is evident from Eq. (2.6), in turn, that the Fourier component of the scalar potential on the surface of the film is related to $\Phi(\mathbf{k},\omega)$ by the equation

$$V(y=0,\mathbf{k},\omega)=-i\omega\frac{\phi_0}{2\pi c}\Phi(\mathbf{k},\omega). \tag{A10}$$

Taking relations (A2) and (A10) into account, we can write the solution of the differential equations (A3) and (A4) in the region $y\geqslant 0$ in the form

$$A_x(y,\mathbf{k},\omega)=a_x\exp(-\psi y)+\frac{c}{\omega}k_xV(y,\mathbf{k},\omega), \tag{A11}$$

$$A_y(y,\mathbf{k},\omega)=\frac{i\mathbf{k}\cdot\mathbf{a}}{\psi}\exp(-\psi y)+i\frac{c}{\omega}kV(y,\mathbf{k},\omega), \tag{A12}$$

$$A_z=(y,\mathbf{k},\omega)=a_z\exp(-\psi y)+\frac{c}{\omega}k_zV(y,\mathbf{k},\omega), \tag{A13}$$

$$V(y,\mathbf{k},\omega)=-i\omega\frac{\phi_0}{2\pi c}\Phi(\mathbf{k},\omega)\exp(-k\psi), \tag{A14}$$

where the function ψ is described by Eq. (2.16). The sign of the imaginary part of the function ψ is chosen so that for $\omega^2>c^2k^2$ the vector potential $\mathbf{A}$ describes the emission of electromagnetic waves into vacuum. We also assume that the scalar potential (and for $\omega^2<c^2k^2$ the vector potential) tends to zero far from the film. Substituting the solutions (A11)–(A14) into the boundary conditions (A5)–(A7), we find

$$a_x=\frac{\phi_0}{2\pi}\frac{\varphi(k_z,\omega)}{1+2\lambda_e\psi}\left(1+\frac{2\lambda_e}{G}k_x^2\right), \tag{A15}$$

$$a_z=-\frac{\phi_0}{2\pi}\frac{\varphi(k_z,\omega)}{1+2\lambda_e\psi}\frac{2\lambda_e}{G}k_xk_z, \tag{A16}$$

$$\Phi(\mathbf{k},\omega)=i\varphi(k_z,\omega)k_x/kG, \tag{A17}$$

where G is defined by Eq. (2.15). Equations (A11)–(A17) express the coupling of the electromagnetic potentials with the phase difference at the tunnel junction [see Eqs. (2.10)–(2.13)].

APPENDIX B:

In general the real part of the Fourier component of the kernel $Q'(k_z,\omega)$ is equal to the sum of four integrals I_n ($n=1,2,3,4$):

$$Q'(k_z,\omega)=I_1+I_2+I_3+I_4, \tag{B1}$$

which have the form

$$I_1=\frac{1}{\pi}\int_{\kappa\eta(\kappa^2)}^{\infty}dk_x\frac{k_z^2}{k^2}\frac{\sqrt{k^2-\omega^2/c^2}}{1+2\lambda_e\sqrt{k^2-\omega^2/c^2}}, \tag{B2}$$

$$I_2=\frac{\omega^2}{\pi c^2}\int_{\kappa\eta(\kappa^2)}^{1/2d}dk_x\frac{k_x^2}{k^2}\frac{1}{2\lambda_e\omega^2/c^2-\sqrt{k^2-\omega^2/c^2}}, \tag{B3}$$

$$I_3=\frac{2}{\pi}\lambda_ek_z^2\eta(\kappa^2)\int_0^{\kappa}\frac{dk_x}{k^2}\frac{\omega^2/c^2-k^2}{1+4\lambda_e^2(\omega^2/c^2-k^2)}, \tag{B4}$$

$$I_4=\frac{2}{\pi}\lambda_e\frac{\omega^4}{c^2}\eta(\kappa^2)\int_0^{\kappa}dk_x\frac{k_x^2}{k^2}\frac{1}{\omega^2/c^2-k^2+4\lambda_e^4\omega^4/c^2}, \tag{B5}$$

where $\kappa^2\equiv\omega^2/c^2-k_z^2$. Having assumed in the initial equations that the scale of the field variation is large in comparison with the thickness of the tunnel junction $2d$, we can eliminate the logarithmic divergence in Eq. (B3) by setting the upper limit of integration equal to $1/2d$. The integrals can be evaluated in quadratures, but the results are fairly involved. We confine the analysis to limiting values of the integrals, which are needed to set forth the basic material underlying the discussion of conditions whereby the parameter $2\lambda_e|\omega|/c$ is smaller than unity. The integrals I_n have a particularly simple form in the limit $|\omega|\ll c|k_z|$, when

$$I_1\simeq\frac{k_z^2}{\pi}Q(2\lambda_e|k_z|),\quad I_2\simeq-\frac{\omega^2}{\pi c^2}\ln\frac{1}{d|k_z|},\quad I_3=I_4=0, \tag{B6}$$

where the function $Q(x)$ is described by Eq. (2.28). In the opposite, high-frequency limit $|\omega|\gg c|k_z|$ we can obtain the following relations from Eqs. (B2)–(B5):

$$I_1 \simeq \frac{k_z^2}{\pi}\ln\frac{c}{\lambda_e|\omega|}, \quad I_2 \simeq -\frac{\omega^2}{\pi c^2}\ln\frac{c}{d|\omega|},$$

$$I_3 \simeq \frac{2}{\pi}\lambda_e|k_z|\frac{\omega^2}{c^2}\arctan\sqrt{\frac{\omega^2}{c^2k_z^2}-1},$$

$$I_4 \simeq \frac{2}{\pi}\lambda_e\frac{|\omega|^3}{c^3}\ln\frac{c}{\lambda_e|\omega|} - I_3. \tag{B7}$$

Equations (B6) and (B7) are sufficiently accurate for basic calculations.

APPENDIX C:

To compute integrals of the form

$$I = \int_0^\kappa dk\, f(k)\exp[iF(k)], \tag{C1}$$

where the function $F(k)$ is described by the expression

$$F(k) = xk + y\sqrt{\kappa^2-k^2}, \tag{C2}$$

we use the steepest descent method. If $y/x=\tan\xi>0$, the function $F(k)$ has a maximum at

$$k_m = \kappa\cos\xi. \tag{C3}$$

The following expansion is valid in the vicinity of the maximum:

$$F(k) = \kappa r_\perp - \frac{r_\perp}{2\kappa\sin^2\xi}(k-\kappa\cos\xi)^2 + \ldots, \tag{C4}$$

where $r_\perp=\sqrt{x^2+y^2}$ is the distance from the Josephson junction to the observation point. Under the conditions

$$\tan(\xi/2) \gg \sqrt{2/\kappa r_\perp}, \quad \tan\xi \ll \sqrt{\kappa r_\perp}/2, \tag{C5}$$

which hold at sufficiently large distances, the limits of integration in Eq. (C1) can be set equal to $\pm\infty$. Assuming then that the function $f(k)$ varies only slightly in the narrow interval

$$\kappa\cos\xi - \sqrt{\frac{2\kappa}{r_\perp}}\sin\xi \leqslant k \leqslant \kappa\cos\xi + \sqrt{\frac{2\kappa}{r_\perp}}\sin\xi, \tag{C6}$$

from Eqs. (C1) and (C4) we obtain

$$I \approx \sqrt{\frac{\pi\kappa}{r_\perp}}f(\kappa\cos\xi)\sin\xi\exp(i\kappa r_\perp)(1-i). \tag{C7}$$

For $\tan\xi<0$ the function $F(k)$ reaches a maximum with respect to k at the boundaries of the domain of integration. The main contribution to the integral I in this case is from the edge of the domain of integration, and the value of the integral itself decreases $\propto 1/r_\perp$. Corrections of this order are insignificant in regard to the radiation field at large distances.

APPENDIX D:

We now show how radiation flows from the surface of the superconductor into vacuum in the case of homogeneous large-scale states. Assuming that the phase difference depends only on time, from the expressions for the potentials (2.10)–(2.14) we can obtain the field on the surface of the film at $|x|\geqslant d$:

$$\mathbf{H}(x,y=+0,t) = -\frac{\phi_0}{(2\pi)^3c^2}\frac{\partial}{\partial t}\int d\omega dk_x dt' e^{i\omega(t'-t)} \times\cos(xk_x)\frac{d\varphi(t')}{dt'}\frac{1}{\psi-2\lambda_e\omega^2/c^2}\mathbf{e}_z, \tag{D1}$$

$$E_x(x,y=0,t) = \frac{\phi_0}{(2\pi)^3c}\int d\omega dt' dk_x e^{i\omega(t'-t)} \times\cos(xk_x)\frac{d\varphi(t')}{dt'}\frac{1}{G}\frac{G+2\lambda_e k_x^2}{1+2\lambda_e\psi}, \tag{D2}$$

$$E_y(x,y=+0,t) = -\frac{\phi_0}{(2\pi)^3c}\int d\omega dt' dk_x e^{i\omega(t'-t)} \times\sin(xk_x)\frac{d\varphi(t')}{dt'}\frac{k_x}{\psi-2\lambda_e\omega^2/c^2}, \tag{D3}$$

where the function ψ is given by Eq. (2.16), in which it is now required to set $k_z=0$. The magnetic field at the boundary of the tunnel junction with vacuum is determined from Eq. (D1) at $x=0$, and the electric field in the junction is determined from the Josephson equation

$$\mathbf{E}(x,y=0,t) = \frac{\phi_0}{4\pi cd}\frac{d\varphi(t)}{dt}\mathbf{e}_x. \tag{D4}$$

For the nonlinear states (5.10) and (5.19), when the time derivative of the phase difference has the form

$$\frac{d\varphi(t)}{dt} = \sum_{n=0}^{\infty} a_n\cos\Omega_n t,$$

where $\Omega_n=(2n+1)\Omega_v$ for the solution (5.10), and $\Omega_n=2n\Omega_r$ for the solution (5.19), from Eqs. (D1)–(D3) we obtain equations for the field outside the nonsuperconducting layer:

$$\mathbf{H}(x,y=+0,t) = \frac{\phi_0}{(2\pi c)^2}\sum_{n=0}^{\infty} a_n\Omega_n\int dk_x \times\cos(xk_x)\,\mathrm{Im}\left[\frac{e^{i\Omega_n t}}{\psi_n-2\lambda_e\Omega_n^2/c^2}\right]\mathbf{e}_z, \tag{D5}$$

$$E_x(x,y=0,t) = \frac{\phi_0}{(2\pi)^2c}\sum_{n=0}^{\infty} a_n\int dk_x \times\cos(xk_x)\,\mathrm{Re}\left[\frac{e^{i\Omega_n t}}{G_n}\frac{G_n+2\lambda_e k_x^2}{1+2\lambda_e\psi_n}\right], \tag{D6}$$

$$E_y(x,y=+0,t) = -\frac{\phi_0}{(2\pi)^2c}\sum_{n=0}^{\infty} a_n\int dk_x \times\sin(xk_x)k_x\,\mathrm{Re}\left[\frac{e^{i\Omega_n t}}{\psi_n-2\lambda_e\Omega_n^2/c^2}\right], \tag{D7}$$

where

$$\psi_n \equiv \sqrt{k_x^2-\Omega_n^2/c^2}\ \eta(c^2k_x^2-\Omega_n^2) + i\sqrt{\Omega_n^2/c^2-k_x^2}\ \eta(\Omega_n^2-c^2k_x^2),$$

$$G_n \equiv \psi_n - 2\lambda_e\Omega_n^2/c^2.$$

As before [see Eq. (3.3)], we assume that the electromagnetic wavelength in vacuum is greater than the effective London penetration depth of the magnetic field in the film,

$$2\lambda_e\Omega_n/c \ll 1. \tag{D8}$$

Then, ignoring small corrections containing the parameter (D8), from Eqs. (D5)–(D7) we can obtain an equation for the flux density on the surface of the film away from the tunnel junction, averaged over the fundamental period:

$$\mathbf{S} = \frac{r_\perp}{x}\sum_{n=0}^{\infty} S_n \mathbf{e}_x, \tag{D9}$$

where S_n is given by Eq. (5.15) or (5.22). According to Eq. (D9), the flux density tends to zero in the limit $|x|\to\infty$. This means that in the limit (D8) the radiative energy losses from the surface of the superconductor are negligible. On the other hand, under conditions (D8), taking (D4) and (D5) into account, we can write the energy flux density from the tunnel junction at the junction–vacuum interface in the form (for $y=+0$)

$$\mathbf{S} = \frac{\pi r_\perp}{2d}\sum_{n=0}^{\infty} S_n \mathbf{e}_y,$$

Under condition (D8) the total energy lost by the junction from unit length along the z axis per unit time is

$$4d|\mathbf{S}| = 2\pi r_\perp \sum_{n=0}^{\infty} S_n.$$

The integration of Eq. (5.14) or (5.21) over all angles ξ [see Eq. (4.18)] produces the same result.

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

[1] I. O. Kulik and I. K. Yanson, *The Josephson Effect in Superconducting Tunnel Structures* [in Russian], Nauka, Moscow (1970).
[2] K. K. Likharev, *Dynamics of Josephson Junctions and Circuits*, Gordon and Breach, New York (1986).
[3] A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect*, Wiley, New York (1982).
[4] K. L. Ngai, Phys. Rev. **182**, 555 (1969).
[5] M. B. Mineev and V. V. Shmidt, Zh. Éksp. Teor. Fiz. **79**, 893 (1980) [Sov. Phys. JETP **52**, 453 (1980)].
[6] Yu. S. Kivshar and B. A. Malomed, Phys. Rev. B **37**, 9325 (1988).
[7] A. A. Golubov, I. L. Serpuchenko, and A. V. Ustinov, Zh. Éksp. Teor. Fiz. **94**(6), 297 (1988) [Sov. Phys. JETP **67**, 1256 (1988)].
[8] V. P. Silin and A. V. Studenov, Kratk. Soobshch. Fiz., No. 1–2, 71 (1996).
[9] V. V. Kurin and A. V. Yulin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **38**, 287 (1995).
[10] K. N. Ovchinnikov, V. P. Silin, and S. A. Uryupin, Fiz. Met. Metalloved. **83**, 14 (1997).
[11] J. Pearl, Appl. Phys. Lett. **5**, 65 (1964).
[12] Yu. M. Ivanchenko and T. K. Soboleva, JETP Lett. **51**, 114 (1990).
[13] Yu. M. Ivanchenko and T. K. Soboleva, Phys. Lett. A **147**, 65 (1990).
[14] R. G. Mints and I. B. Shapiro, Phys. Rev. B **51**, 3054 (1995).
[15] B. I. Ivlev and N. B. Kopnin, Adv. Phys. **33**, 47 (1984).
[16] Yu. M. Kupriyanov, K. K. Likharev, and A. K. Semenov, Fiz. Nizk. Temp. **2**, 706 (1976) [Sov. J. Low Temp. Phys. **2**, 346 (1976)].
[17] A. Gurevich, Phys. Rev. B **48**, 13857 (1993).
[18] V. P. Silin, Zh. Eksp. Teor. Fiz. **112**, 1396 (1997) [JETP **85**, 760 (1997)].
[19] Z. D. Genchev and V. I. Vas'kiv'skiĭ, Zh. Éksp. Teor. Fiz. **113**, 955 (1998) [JETP **86**, 521 (1998)].
[20] R. G. Mints, J. Low Temp. Phys. **106**, 183 (1997).
[21] A. S. Malishevskiĭ, V. P. Silin, and S. A. Urovnin, Kratk. Soobshch. Fiz., No. 3, 21 (1998).
[22] A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **53**, 2159 (1967) [Sov. Phys. JETP **26**, 1219 (1967)].
[23] K. K. Likharev, V. K. Semenov, and A. B. Zorin, *New Possibilities for Superconductor Electronics* [in Russian], VINITI, Moscow (1988).

Translated by James S. Wood

Magnetic response of a two-dimensional degenerate electron gas in nanostructures with cylindrical symmetry

V. A. Geĭler, V. A. Margulis, and A. V. Shorokhov

N. P. Ogarev Mordovan State University, 430000 Saransk
(Submitted 15 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1450–1462 (April 1999)

An investigation is made of the magnetic response of nanostructures with cylindrical symmetry located in a longitudinal magnetic field. Analytic expressions are obtained for the magnetic moment of the nanostructures, cylinders and bracelets. It is shown that the magnetic moment describes Aharonov–Bohm oscillations. The profile of the oscillations and the position of the oscillation maxima are studied. In the limit $T\to 0$ the curves of the magnetic response as a function of the magnetic field flux contain ''beak''-shaped kinks, and the positions of the points at which these kinks occur are determined. The temperature dependence of the magnetic response is studied and the influence of the spin–magnetic interaction on the magnetic response of the nanostructures is examined. It is shown that this interaction destroys the periodicity of the magnetic response with respect to flux and gives rise to a monotonic term in the response. © *1999 American Institute of Physics.* [S1063-7761(99)02004-1]

1. INTRODUCTION

Over the last few years, the equilibrium and transport properties of electrons in mesoscopic samples have been the focus of intensive experimental and theoretical studies.[1] This is because modern technologies can be used to fabricate perfect nanostructures of various geometries (quantum wells and dots, channels, wires, and rings in heterostructures). Each of these nanostructures has its own unique physical properties. Interesting equilibrium effects occur when a magnetic field is applied to a sample.[8–11]

Note that studies of the magnetic response in various quasi-one-dimensional and quasi-two-dimensional systems can give important information on the parameters of the electron energy spectrum and the lateral confinement potentials in these structures.[1–10] Moreover, a magnetic field applied to the nanostructure can provide additional scope for studying its parameters. This is because a magnetic field can create or enhance the existing lateral confinement in the nanostructure and can also produce hybrid coupling between motion parallel and perpendicular to the field when the field is directed at an angle to the symmetry axis of the system.[4,7]

The equilibrium properties of an electron gas in nanostructures are mainly determined by the electron energy spectrum, which is itself determined by the geometry of the system. In addition to the nanostructures described above, which are fabricated in a planar two-dimensional electron system, investigations of nanosystems in curved layers have also started recently.[11–18] Special procedures (such as lift-off) have recently allowed us to investigate a curved layer of electron gas experimentally.[17] The geometry of a cylinder is the closest to the experimental situation[17] and in particular, the geometry of a carbon nanotube is close to that of a cylinder. The magnetic response of a nanotube at $T=0$ was investigated numerically by Ajiki and Ando.[19] The effective-mass approximation was used to find the electron spectrum for the case where the magnetic field **B** is parallel to the tube axis. Lin and Shung[2] investigated the magnetic response of a nanotube for the same field orientation using the strong-coupling approximation but allowing for spin–magnetic interaction. The latter gives rise to a crescent-shaped singularity (cusp) on the curve giving the magnetization M as a function of the magnetic field flux Φ near integer values of the quantum flux $\Phi_0=e/ch$. In Ref. 2 it was noted that the results for the strong-coupling model are similar to those obtained in the effective-ma ss approximation.[19]

Ovchinnikov *et al.*[18] reported a detailed theoretical analysis of the average magnetic response of various mesoscopic systems, including a bracelet (a cylinder with height of order the Fermi wavelength). The magnetic field was directed normal to the side surface of the cylinder. For the averaging, these authors assumed that the chemical potential has a random correction which is distributed uniformly in an interval on the order of the interlevel spacing. As a result of this averaging, both the de Haas–van Alphen oscillations and the dimensional fluctuations are smeared out. Consequently for $T\gg\hbar\omega_c$ Ovchinnicov *et al.*[18] obtained expressions for the magnetic susceptibility of these mesoscopic systems which are an analog of the Landau formula for the diamagnetic susceptibility.

In view of this reasoning, it is quite important to derive analytic expressions for the magnetic moment in nonplanar two-dimensional nanosystems and to study Aharonov–Bohm oscillations of the magnetic response, as well as determining the temperature dependence of the profile of these oscillations. The main purpose of the present study is to obtain convenient formulas for analytic and numerical investigations, which describe the oscillations of the magnetic response of a two-dimensional degenerate electron gas folded into a cylinder (quantum cylinder), including a cylinder with a short generatrix whose length L is of the order of the Fermi

wavelength of an electron λ_F (quantum bracelet). The topical case of a magnetic field directed along the symmetry axis of the system is studied (a static homogeneous field **B** and a field generated by an infinitely thin Aharonov–Bohm solenoid are both considered). An analysis is made of the influence of the electron spin on the profile of the oscillation curve. As will be shown subsequently, the nontrivial curvature of these two-dimensional systems leads to an interesting new effect: in the limit $T\to 0$ beak-shaped kinks appear on the oscillation curve. The nature of these kinks is similar to that responsible for the kinks on the curve of the undamped current in a quantum magnetic ring.[3]

Finally, we note that the magnetic response is investigated using both a Gibbs canonical distribution (constant number of electrons) and a large Gibbs canonical distribution (constant chemical potential of the gas: $\mu(B)=\text{const}$). For most situations the results obtained using these distributions differ very little. This is because the oscillating component of $\mu(B)$ for a constant number of electrons is very small.[20] For a degenerate electron gas, however, it is more convenient to use the large canonical distribution for the calculations. Thus, we shall subsequently adopt this approach, i.e., we shall assume that $\mu=\text{const}$. In addition, we shall exclusively consider a noninteracting electron gas.

2. MAGNETIC RESPONSE OF A QUANTUM CYLINDER

In the effective-mass approximation the Hamiltonian H of the single-electron spin-zero states for the vector potential **A**, taken in the form $\mathbf{A}=(By/2,-Bx/2,0)$, is written in cylindrical coordinates as

$$H=-\varepsilon\frac{d^2}{d\varphi^2}-\frac{i\hbar\omega_c}{2}\frac{d}{d\varphi}+\frac{m^*\omega_c^2}{8}\rho^2+\frac{p^2}{2m^*}, \tag{1}$$

where $\omega_c=eB/m^*c$ is the cyclotron frequency, φ is the polar angle, m^* is the effective electron mass, p is the momentum in the direction of the cylinder axis, and $\varepsilon=\hbar^2/2m^*\rho^2$ is the dimensional confinement energy.

The spectrum of the Hamiltonian has the form

$$\varepsilon_{mp}=\varepsilon\left(m+\frac{\Phi}{\Phi_0}\right)^2+\frac{p^2}{2m^*}. \tag{2}$$

Here the magnetic quantum number m has the values $m=0,\ \pm1,\ \pm2,\ \ldots$, and $\Phi=\pi\rho^2B$ is the flux of the field **B** through the cross section of the cylinder. Using the standard expression for the thermodynamic potential Ω (Ref. 21), we obtain in our case

$$\Omega=-\frac{TL}{2\pi\hbar}\sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty}dp\ln\left[1+\exp\left(\frac{\mu-\varepsilon_{mp}}{T}\right)\right], \tag{3}$$

where L is the cylinder length. From expression (3) we find the magnetic moment using the formula $M=-(\partial\Omega/\partial B)_{\mu/T}$ and then

$$-\frac{M}{\mu_B}=\frac{Lm_0}{\pi\hbar m^*}\sum_{m=-\infty}^{\infty}\int_0^{\infty}\frac{(m+\Phi/\Phi_0)\,dp}{1+\exp[(\varepsilon_{mp}-\mu)/T]}, \tag{4}$$

where m_0 is the free electron mass and μ_B is the Bohr magneton.

For the following analysis it is convenient to expand the magnetic moment of the cylinder as a Fourier series using a Poisson summation formula. After some simple but fairly lengthy transformations, we obtain

$$-\frac{M}{\mu_B}=\sum_{n=1}^{\infty}C_n(T)\sin\left(2\pi n\frac{\Phi}{\Phi_0}\right), \tag{5}$$

where the Fourier coefficients $C_n(T)$ have the form

$$C_n(T)=\frac{Lm_0}{\pi^3\hbar m^*}\int_0^{\infty}dz\,z\sin(nz)\times\int_0^{\infty}dp\left\{1+\exp\left[\frac{1}{T}\left(\frac{\varepsilon z^2}{4\pi^2}+\frac{p^2}{2m^*}-\mu\right)\right]\right\}^{-1}. \tag{6}$$

We introduce the new variables $x=z\sqrt{\varepsilon}/2\pi$, $y=p/\sqrt{2m^*}$, and convert in Eq. (6) to polar coordinates (r,ψ) in the xy plane. For $C_n(T)$ we then obtain

$$C_n(T)=\frac{4L\sqrt{2m^*}m_0}{\hbar m^*\varepsilon\pi}\int_0^{\infty}dr\,r^2\left[1+\exp\left(\frac{r^2-\mu}{T}\right)\right]^{-1}\times\int_0^{\pi/2}\sin\left(\frac{2\pi nr\cos\psi}{\sqrt{\varepsilon}}\right)\cos\psi\,d\psi. \tag{7}$$

Using an integral representation for the Bessel function J_1 (Ref. 22),

$$\int_0^{\pi/2}\sin\left(\frac{2\pi nr\cos\psi}{\sqrt{\varepsilon}}\right)\cos\psi\,d\psi=\frac{\pi}{2}J_1\left(\frac{2\pi nr}{\sqrt{\varepsilon}}\right), \tag{8}$$

we obtain

$$C_n(T)=\frac{2L\sqrt{2m^*\varepsilon}m_0}{\hbar m^*}\int_0^{\infty}\frac{x^2J_1(2\pi nx)dx}{1+\exp[(\varepsilon x^2-\mu)/T]}. \tag{9}$$

It is deduced from Eqs. (9) and (5) that the magnetic moment of the quantum cylinder is an oscillating function of the flux with a period equal to a flux quantum.

For a quantitative analysis of the nature of the oscillations we consider the case $T=0$. Then

$$C_n(0)=\frac{2L\sqrt{2m^*\varepsilon}m_0}{\hbar m^*}\int_0^{\sqrt{\mu/\varepsilon}}x^2J_1(2\pi nx)dx. \tag{10}$$

Using the formula[22]

$$\int_0^1x^{\nu+1}J_\nu(ax)\,dx=a^{-1}J_{\nu+1}(a), \tag{11}$$

we obtain

$$C_n(0)=\frac{L\mu\sqrt{2m^*}m_0}{\pi\hbar m^*\sqrt{\varepsilon}}\frac{J_2(2\pi n\sqrt{\mu/\varepsilon})}{n}. \tag{12}$$

In real situations we have $\mu\gg\varepsilon$. Using the asymptotic form of the Bessel function $J_2(x)$ for large values of the argument[22] we obtain the following estimate from Eq. (12)

$$C_n(0)=\frac{L\mu\sqrt{2m^*}m_0}{\pi^2\hbar m^*\sqrt{\varepsilon}}\frac{1}{n}\left[\left(\frac{\sqrt{\varepsilon}}{\pi^2 n\sqrt{\mu}}\right)^{1/2}\times\cos\left(2\pi n\sqrt{\frac{\mu}{\varepsilon}}-\frac{5\pi}{4}\right)+o\left(\frac{\sqrt{\varepsilon}}{2\pi n\sqrt{\mu}}\right)\right]. \quad (13)$$

Equation (13) then yields an estimate for $M(T=0)$

$$\frac{M(T=0)}{\mu_B}=\frac{\sqrt{2}Lm_0}{\pi\hbar\sqrt{m^*}}\left(\frac{\mu^3}{\varepsilon}\right)^{1/4}\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\times\sin\left(2\pi n\frac{\Phi}{\Phi_0}\right)\cos\left(2\pi n\sqrt{\frac{\mu}{\varepsilon}}-\frac{\pi}{4}\right). \quad (14)$$

Formula (14) yields an important statement for the following analysis: because of the periodicity of the functions appearing in the Fourier series the magnetic moment depends only on the fractional component of Φ/Φ_0 and $\sqrt{\mu/\varepsilon}$ in Eq. (14). We shall denote these by ξ and η, respectively ($0\leqslant\eta$, $\xi<1$). Then we have

$$\frac{M(T=0)}{\mu_B}=\frac{\sqrt{2}Lm_0}{\pi^2\hbar\sqrt{m^*}}\left(\frac{\mu^3}{\varepsilon}\right)^{1/4}\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\times\sin(2\pi n\xi)\cos\left(2\pi n\eta-\frac{\pi}{4}\right). \quad (15)$$

Expression (15) indicates that it is sufficient to study the oscillations of the moment in the region $0\leqslant\xi<1$, $0\leqslant\eta<1$ since, as Φ/Φ_0 varies further, the pattern is repeated for each separate segment of variation. The sum of the series in Eq. (15) may be expressed in terms of generalized ζ-functions, using the Hurwitz formula[23]

$$\sum_{n=1}^{\infty}\frac{1}{n^{3/2}}\begin{Bmatrix}\sin(2\pi nx)\\ \cos(2\pi nx)\end{Bmatrix}=2\pi[\zeta(-1/2,x)\pm\zeta(-1/2,1-x)], \quad (16)$$

where $0<x\leqslant 1$.

We first note the symmetry properties of the graph of $M(\Phi/\Phi_0)$. Expression (15) indicates that $M(\xi,\eta)=-M(1-\xi,\eta)$ holds for any $0\leqslant\eta<1$. Thus, the graph is antisymmetric relative to the axis passing through the point $\xi=1/2$ perpendicular to the ordinate. In addition we have $M(\xi+1/2,\eta-1/2)=M(\xi,\eta)$. Consequently the case $1/2\leqslant\eta<1$ is reduced to $\eta<1/2$ by shifting the graph of $M(\Phi/\Phi_0)$ along the abscissa by half a flux quantum. On account of this symmetry, we shall only consider the region $\xi,\eta\leqslant 1/2$, i.e., half the period of the function $M(\Phi/\Phi_0)$. We immediately note that for integer values of $\sqrt{\mu/\varepsilon}$ this half period has only one extremum (maximum or minimum depending on whether $\eta<1/2$ or $\eta>1/2$). For $\eta<1/2$ formulas (15) and (16) give

$$\frac{M(\xi,\eta)}{4\pi A\mu_B}=\zeta(-1/2,\xi+\eta)-\zeta(-1/2,1-\xi+\eta),\quad \xi\geqslant\eta, \quad (17a)$$

$$\frac{M(\xi,\eta)}{4\pi A\mu_B}=\zeta(-1/2,\xi+\eta)-\zeta(-1/2,\eta-\xi),\quad \eta>\xi, \quad (17b)$$

where $Lm_0/2\pi^2\hbar\sqrt{m^*}(\mu^3/\varepsilon)^{1/4}\equiv A$.

Expression (17) indicates that the critical point of the graph of $M(\xi)$ on the interval $0<\xi<1/2$ is at $\xi=\eta$. We shall analyze the behavior of the graph near this point using the shift formula for the generalized ζ-function[23]

$$\zeta(s,x)=\zeta(s,1+x)+x^{-s}. \quad (18)$$

Expression (17b) is then written as

$$\frac{M(\xi,\eta)}{4\pi A\mu_B}=\zeta(-1/2,\xi+\eta)-\zeta(-1/2,1-\xi+\eta)-\sqrt{\eta-\xi},\quad \eta>\xi. \quad (19)$$

The functions $\zeta(-1/2,x)$ are continuous, as is deduced from Eqs. (15) and (16) since the corresponding Fourier series converge uniformly. However, a comparison of Eqs. (17a) and (19) shows that at the point $\eta=\xi$ the graph of $M(\xi,\eta)$ has a kink caused by the presence of the third term in formula (19). Moreover, this term may give an additional zero of the function $M(\xi,\eta)$ in the range $0<\xi<1/2$. In this case, two extrema (a maximum and a minimum) are observed at half the period of the function.

Figure 1 shows graphs plotted using formula (15). These graphs are consistent with the analytic results presented above on the behavior of the magnetic moment.

For $T\neq 0$ the kinks on the graphs are smoothed, but the general oscillation pattern is still retained (Fig. 2).

3. MAGNETIC RESPONSE OF A QUANTUM BRACELET

We shall consider a cylinder with a short generatrix, i.e., a quantum bracelet. When the cylinder length is comparable with the Fermi wavelength of an electron, we must take into account the quantization of the particle motion along the z axis. We shall take the model of an infinitely deep potential well as a model of the confinement potential along this axis. The electron spectrum for this case is then written in the form

$$E_{nm}=\varepsilon_m+\varepsilon_n,\quad m=0,\pm1,\pm2,\ldots,\quad n=1,2,\ldots, \quad (20)$$

where $\varepsilon_m=a(2\pi m+2\pi\Phi/\Phi_0)$ and $\varepsilon_n=b(2\pi n)^2$. The dimensional confinement energies are $a=\hbar^2/8m^*\pi^2\rho^2$ and $b=\hbar^2/8m^*L^2$. The bracelet length is L and its radius ρ.

Using the thermodynamic potential of the electron gas in the bracelet,

$$\Omega=-T\sum_{n,m}\ln\left[1+\exp\left(\frac{\mu-E_{nm}}{T}\right)\right], \quad (21)$$

we obtain the following expression for the magnetic response of the bracelet

$$-\frac{M}{\mu_B}=\frac{m_0}{m^*}\sum_{m=-\infty}^{\infty}\sum_{n=1}^{\infty}\frac{m+\Phi/\Phi_0}{1+\exp[(E_{nm}-\mu)/T]}. \quad (22)$$

Expression (22) can be conveniently rewritten as

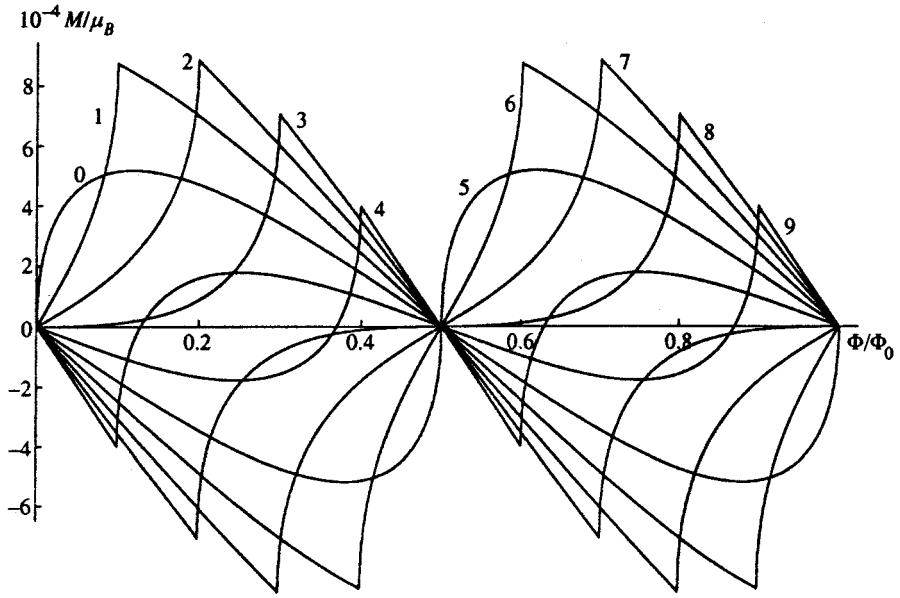

FIG. 1. Magnetic response of a quantum cylinder as a function of the magnetic field flux. The curve number N corresponds to $\eta=0,N$, where N has values $N=0,\ 1,\ \ldots,\ 9$ (for example $\eta=0,\ 0,\ 1,\ \ldots$.

$$-\frac{M}{\mu_B}=\frac{1}{2}\frac{m_0}{m^*}\sum_{m,n=-\infty}^{\infty}\frac{m+\Phi/\Phi_0}{1+\exp[(E_{nm}-\mu)/T]}$$
$$-\frac{1}{2}\frac{m_0}{m^*}\sum_{m=-\infty}^{\infty}\frac{m+\Phi/\Phi_0}{1+\exp[(E_{0m}-\mu)/T]}. \qquad (23)$$

The second sum in expression (23) is half the magnetic moment of the quantum ring $M_{\rm ring}/\mu_B$.

Cheung *et al.*[3] studied an undamped current J in an isolated quantum ring. This current was calculated using the free energy F, i.e., assuming that the number of particles is constant. The value of J is proportional to the magnetic moment of the ring. Since the expression for $M_{\rm ring}$ appears in the formula for the magnetic moment of the bracelet (which is assumed to be the constant component of a large system, for instance, the bracelet is located in the quasi-one-dimensional layer of a heterostructure), we give the expression for $M_{\rm ring}$ for completeness.

Assuming that the chemical potential μ is constant, we obtain the following expression for the expansion of $M_{\rm ring}$ as a Fourier series

$$-\frac{M_{\rm ring}}{\mu_B}=\sum_{n=1}^{\infty}C_n\sin\left(2\pi n\frac{\Phi}{\Phi_0}\right), \qquad (24)$$

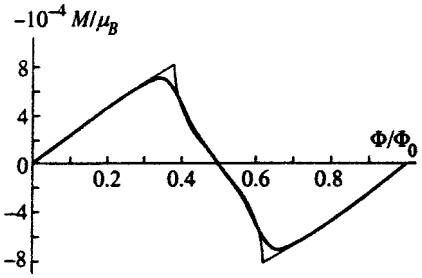

FIG. 2. Temperature smoothing of the magnetic response curve of a quantum cylinder. The fine line corresponds to $T=0$ and the heavy line corresponds to $T=10$ K, $\eta=0.6$, $\rho=3.58\times10^{-6}$, and $L=4.37\times10^{-4}$.

where the Fourier coefficients are given by

$$C_n(T)=\frac{4m_0}{m^*}\int_0^{\infty}\frac{x\sin(2\pi nx)dx}{1+\exp[(\varepsilon x^2-\mu)/T]}. \qquad (25)$$

In the low-temperature limit ($T\to0$) Eq. (25) gives

$$C_n(0)=\frac{m_0}{m^*}\left[\frac{1}{\pi^2n^2}\sin\left(2\pi n\sqrt{\frac{\mu}{\varepsilon}}\right)\right.$$
$$\left.-\sqrt{\frac{\mu}{\varepsilon}}\frac{2}{\pi n}\cos\left(2\pi n\sqrt{\frac{\mu}{\varepsilon}}\right)\right]. \qquad (26)$$

The series (24) may be summed using Eq. (26). Denoting the integer part of $\sqrt{\mu/\varepsilon}$ by N, we obtain for $\xi+\eta<1$

$$-\frac{M_{\rm ring}}{\mu_B}=\frac{m_0}{m^*}\begin{cases}(2N+1)\xi, & \xi\leqslant\eta,\\ 2N(\xi-1/2), & \eta<\xi,\end{cases} \qquad (27)$$

and for $\xi+\eta\geqslant1$

$$-\frac{M_{\rm ring}}{\mu_B}=\frac{m_0}{m^*}\begin{cases}(2N+2)(\xi-1/2), & \xi\leqslant\eta,\\ (2N+1)(\xi-1), & \eta<\xi.\end{cases} \qquad (28)$$

Expressions (27) and (28) indicate that the magnetic response of the ring describes sawtooth oscillations at $T=0$. The amplitude of these oscillations is proportional to $\sqrt{\mu/\varepsilon}$ and except for the case $\eta=1/2$, two sawtooth maxima of different height are obtained per period. The width of one of these is $|1-2\eta|$ and the other is $1-|1-2\eta|$.

A one-dimensional quantum ring is a limiting case of two-dimensional structures, i.e., a cylinder or a bracelet.

We now write expression (23) in the form

$$-\frac{M}{\mu_B}=\frac{M_{\rm ring}}{\mu_B}$$
$$+\frac{1}{4}\left(\frac{m_0}{m^*}\right)\sum_{m,n=-\infty}^{\infty}\frac{2\pi(m+\Phi/\Phi_0)}{1+\exp[(E_{nm}-\mu)/T]}. \qquad (29)$$

We then expand the magnetic moment of the bracelet as a Fourier series, for which we again use the Poisson formula

$$\sum_{n=-\infty}^{\infty}\varphi(2\pi n+t)=\frac{1}{2\pi}\sum_{l=-\infty}^{\infty}e^{ilt}\int_{-\infty}^{\infty}\varphi(x)e^{-ilx}\,dx. \tag{30}$$

After some fairly lengthy but simple transformations we obtain

$$-\frac{M}{\mu_B}=\frac{M_{\rm ring}}{\mu_B}+\sum_{k=1}^{\infty}C_k(T)\sin(2\pi k\xi), \tag{31}$$

where the Fourier coefficients are

$$C_k(T)=-\frac{1}{2\pi^3}\frac{m_0}{m^*}\sum_{l=-\infty}^{\infty}\int_0^{\infty}dx\cos(lx)\times\int_0^{\infty}\frac{y\sin(ky)dy}{1+\exp[(bx^2+ay^2-\mu)/T]}. \tag{32}$$

Introducing the polar coordinates using the formulas $x=(r/b)\cos\varphi$ and $y=(r/a)\sin\varphi$, we then obtain from Eq. (32)

$$C_k(T)=-\frac{1}{4\pi^2}\frac{m_0}{m^*}\frac{k}{\sqrt{ba^3}}\sum_{l=-\infty}^{\infty}\frac{1}{\sqrt{k^2/a+l^2/b}}\times\int_0^{\infty}\frac{r^2J_1(r\sqrt{k^2/a+l^2/b})\,dr}{1+\exp[(r^2-\mu)/T]}. \tag{33}$$

We shall next analyze the $T=0$ case. The integral in Eq. (33) is easily calculated at $T=0$:

$$\int_0^{\sqrt{\mu}}r^2J_1\left(r\sqrt{\frac{k^2}{a}+\frac{l^2}{b}}\right)dr=\frac{\mu}{\sqrt{k^2/a+l^2/b}}J_2\left(\sqrt{\mu\left(\frac{k^2}{a}+\frac{l^2}{b}\right)}\right). \tag{34}$$

Using the asymptotic form of the Bessel functions for large arguments ($\mu/a,\mu/b\gg1$), we obtain

$$C_k(0)=\frac{1}{4\pi^2}\frac{m_0}{m^*}\sqrt{\frac{2}{\pi}}\left(\frac{\mu^3}{b^2}\right)^{1/4}\left\{\frac{1}{a^{1/4}}\frac{1}{k^{3/2}}\cos\left(\sqrt{\frac{\mu}{\varepsilon}}k-\frac{\pi}{4}\right)+\frac{2k}{a^{5/4}}\sum_{l=1}^{\infty}\frac{\cos[\sqrt{\mu(k^2/a+l^2/b)}-\pi/4]}{(k^2/a+l^2/b)^{5/4}}\right\}. \tag{35}$$

Substituting Eq. (35) into Eq. (31) we then obtain the following formula for the magnetic response of the bracelet

$$-\frac{M(T=0)}{\mu_B}=\frac{M_{\rm ring}(T=0)}{2\mu_B}+\frac{M_{\rm cyl}(T=0)}{2\mu_B}+\frac{4}{(2\pi)^{5/2}}\frac{m_0}{m^*}\left(\frac{\mu^3}{a^6b^2}\right)^{1/4}\sum_{k=1}^{\infty}k\sin\left(2\pi k\frac{\Phi}{\Phi_0}\right)\times\sum_{l=1}^{\infty}\frac{\cos[\sqrt{\mu(k^2/a+l^2/b)}-\pi/4]}{(k^2/a+l^2/b)^{5/4}}, \tag{36}$$

where $M_{\rm cyl}$ is the magnetic response of a cylinder (19) obtained in Sec. 2.

Plots of formulas (31) and (33) are shown in Fig. 2. It can be seen that $M(T)$ goes to zero at points where Φ/Φ_0 is a half-integer (as for a cylinder and a quantum ring). Hence $M(T)$ is a periodic function of the magnetic field flux with a period equal to the flux quantum.

4. TEMPERATURE DEPENDENCE OF THE MAGNETIC RESPONSE

We shall now consider the temperature corrections to the magnetic response, confining our analysis to the case of a highly degenerate gas in the limit $\mu/T\gg1$. In this case, the integral in the expression for the magnetic response of a cylinder (9) and a bracelet (33) may be estimated. We first express $P(T)$, which satisfies

$$P(T)=\int_0^{\infty}\frac{x^2J_1(\alpha x)dx}{1+\exp[(x^2-\mu)/T]},$$

in the form

$$P(T)=\frac{1}{2}\int_0^{\infty}\varphi(x)\left(-\frac{\partial f}{\partial x}\right)dx, \tag{37}$$

where

$$\varphi(x)\simeq2xJ_2(\alpha\sqrt{x})/\alpha,\quad f(x)=1+\exp[(x-\mu)/T]^{-1}.$$

Now $P(T)$ can easily be estimated using the formula[24]

$$\int_0^{\infty}\varphi(x)\left(-\frac{\partial f}{\partial x}\right)dx\simeq\varphi(\mu)+\frac{\pi^2T^2}{6}\varphi''(\mu). \tag{38}$$

Equation (38) gives the estimate

$$P(T)\simeq P(0)+\frac{\alpha\pi^2T^2}{12}J_0(\alpha\sqrt{\mu}). \tag{39}$$

Using Eq. (39) we can find the Fourier coefficients $C_n(T)$ of the magnetic response of a cylinder

$$C_n(T)\simeq C_n(0)+\frac{\pi^3LT^2\sqrt{2m^*\varepsilon}m_0}{3\hbar m^*\varepsilon^2}J_0\left(2\pi n\sqrt{\frac{\mu}{\varepsilon}}\right). \tag{40}$$

The magnetic moment is then given by

$$\frac{M_{\rm cyl}(T)}{\mu_B}\simeq\frac{M_{\rm cyl}(T=0)}{\mu_B}+\frac{\pi^3LT^2\sqrt{2m^*\varepsilon}m_0}{3\hbar m^*\varepsilon^2}\times\sum_{n=1}^{\infty}\sin\left(2\pi n\frac{\Phi}{\Phi_0}\right)J_0\left(2\pi n\sqrt{\frac{\mu}{\varepsilon}}\right). \tag{41}$$

The Fourier coefficients for a bracelet have the form

$$C_n(T)\simeq C_n(0)+\frac{1}{48\pi^2}\frac{m_0}{m^*}\frac{\pi^2T^2n}{\sqrt{a^3b}}\times\sum_{l=-\infty}^{\infty}J_0\left(\sqrt{\mu\left(\frac{k^2}{a}+\frac{l^2}{b}\right)}\right), \tag{42}$$

and its magnetic moment is written as

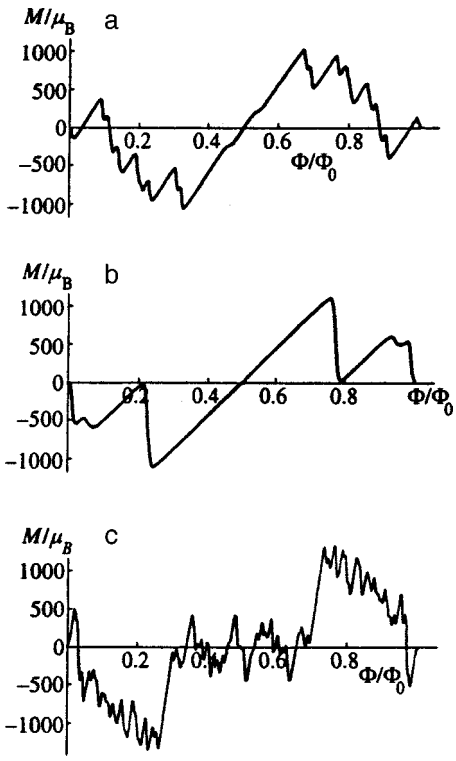

FIG. 3. Magnetic response of a quantum bracelet as a function of the magnetic field flux at $T=1$ K, $\mu=1.6\times10^{-12}$ erg: a — $\rho=2\times10^{-6}$ cm, $L=2\times10^{-6}$ cm; b — $\rho=7\times10^{-6}$ cm, $L=8\times10^{-7}$ cm, c — $\rho=4.7\times10^{-7}$ cm, $L=4.61\times10^{-6}$ cm.

$$\frac{M_{\rm brac}(T)}{\mu_B}=\frac{M_{\rm brac}(T=0)}{\mu_B}+\frac{1}{48}\frac{m_0}{m^*}\frac{T^2}{\sqrt{a^3b}}\times\sum_{k=1}^{\infty}k\sin\left(2\pi k\frac{\Phi}{\Phi_0}\right)\sum_{l=-\infty}^{\infty}J_0\left(\sqrt{\mu\left(\frac{k^2}{a}+\frac{l^2}{b}\right)}\right). \tag{43}$$

The temperature corrections to formulas (41) and (43) smooth the kinks on the curves giving the magnetic moment of a cylinder and a bracelet as a function of the magnetic field flux (Figs. 2 and 3).

5. DISCUSSION OF RESULTS

As is well known, the magnetic moment of a degenerate electron gas in a quantizing magnetic field describes oscillations as the magnetic field varies (de Haas–van Alphen effect). The physical principle of this effect is well-known: the density of electronic states changes abruptly whenever the levels of the electron energy spectrum intersect the level of the chemical potential μ of the gas as the magnetic field varies. Moreover, the amplitude of the oscillation maxima and their position on the curve $M(B)$ also depend on μ.

Note that the set of possible projections of the quasiclassical electron trajectories on the plane perpendicular to the magnetic field is determined by the Fermi energy for $T\to0$ and each energy level has its own projection of this trajectory. The energy levels and consequently the trajectories depend on the value of B and vary as B varies. However, this variation can only take place in such a way that the magnetic field flux across the area occupied by the projection of the trajectory changes by an whole number of flux quanta Φ_0.

With some modifications, the pattern of the de Haas–van Alphen effect described above is also observed for various low-dimension systems.[3–8] However, the results obtained in Secs. 2–4 suggest that in nanostructures exhibiting cylindrical symmetry (quantum ring, bracelet, quantum cylinder), the situation is completely different and although, as before, the amplitudes of the oscillation maxima on the curve $M(B)$ will depend on the chemical potential, the positions of these maxima are determined by the relationships between ξ and η. Moreover, the period of the oscillations does not depend on μ and is equal to the flux quantum (Aharonov–Bohm oscillations). This can be attributed to the purely geometric fact that for any electron energy the projection of its quasiclassical trajectory on the plane perpendicular to the magnetic field is fixed and coincides with the circumference of the cylinder base.

In all the nanostructures considered in the present study the Fourier coefficients do not depend on the magnetic field which distinguishes nanostructures with cylindrical symmetry from ordinary three-dimensional samples[21] or nanostructures possessing no such symmetry.[4–8] We further note that in all cases, the curve of the magnetic response as a function of the field flux Φ has kinks. Exceptions for a cylinder are the cases $\eta=0$ and also $\eta=1/2$ when no kinks are observed. The singularities of the curve for a bracelet incorporate all the singularities of a cylinder and a quantum ring, as is deduced from Eq. (36).

It is interesting to note that the incorporation of an Aharonov–Bohm flux Φ_{AB} (a magnetic field flux created by a thin, ideally infinitely long solenoid positioned along the symmetry axis of the nanostructure) preserves the oscillation pattern for all the cases studied. This flux merely causes a general shift of the magnetic response curve $M(\Phi)$ by Φ_{AB}. This behavior of the magnetic response occurs because the Fourier coefficients do not depend on the magnetic field.

We also note that for all the nanostructures considered the magnetic response has no monotonic component.

These results neglect the electron spin. As a result of allowing for the spin, the term $(1/2)\sigma g\mu_B B$ is added to the electron spectrum, where $\sigma=\pm1$ and g is the electron g-factor. Calculations similar to those made in Secs. 2–4 yield for a quantum ring

$$M_{\rm ring}=\sum_{\sigma=\pm1}(M_\sigma^{\rm mon}+M_\sigma^{\rm osc}).$$

The monotonic component of the ring response $M_\sigma^{\rm mon}$ has the form

$$-\frac{M_\sigma^{\rm mon}}{\mu_B}=\frac{\sigma g}{2}\int_{-\infty}^{\infty}\left\{1+\exp\left[\frac{\varepsilon[x^2+\sigma g(m^*/m_0)(\Phi/\Phi_0)]-\mu}{T}\right]\right\}^{-1}dx, \tag{44}$$

and the oscillating component of the response is described by

$$-\frac{M_\sigma^{\rm osc}}{\mu_B}=\sum_{k=1}^{\infty} a_{k\sigma}\sin\left[2\pi k\frac{\Phi+\Phi_{AB}}{\Phi_0}\right], \tag{45}$$

where the Fourier coefficients are

$$a_{k\sigma}\left(T,\frac{\Phi}{\Phi_0}\right)$$

$$=\frac{4m_0}{m^*}\int_0^\infty \frac{(x+m^*g\sigma/2m_0)\sin(2\pi kx)\,dx}{1+\exp\{(\varepsilon[x^2+\sigma g(m^*/m_0)(\Phi/\Phi_0)]-\mu)/T\}}. \tag{46}$$

Equations (44)–(46) indicate that in the general case where there is an Aharonov–Bohm flux, allowance for the spin causes the following changes in the pattern of the effect: the magnetic response has a monotonic component that depends nonlinearly on the flux Φ and does not depend on the Aharonov–Bohm flux Φ_{AB}, and the Fourier coefficients $a_{k\sigma}$ become nonlinearly dependent on the flux Φ but also do not depend on Φ_{AB}. Thus, as in the case where the spin is neglected, the flux Φ_{AB} only shifts the oscillation pattern. It is important to note that allowance for the spin–magnetic interaction destroys the periodicity of the magnetic response of the ring because the coefficients $a_{k\sigma}$ depend on the flux Φ and the monotonic component of the response depends on the flux.

Similar changes in the oscillation pattern occur in a cylinder and consequently in a bracelet. The magnetic response of a cylinder allowing for spin–magnetic interaction is given by

$$M=\sum_\sigma (M_\sigma^{(1)}+M_\sigma^{(2)}),$$

where

$$\frac{M_\sigma^{(1)}}{\mu_B}=\frac{Lm_0}{\pi m^*\hbar}\sum_{m=-\infty}^{\infty}\int_0^\infty \frac{[m+(\Phi+\Phi_{AB})/\Phi_0]\,dp}{1+\exp[(\varepsilon_{mp}+g\sigma\mu_BB/2-\mu)/T]}, \tag{47}$$

$$\frac{M_\sigma^{(2)}}{\mu_B}=\frac{Lg\sigma}{2\pi\hbar}\sum_{m=-\infty}^{\infty}\int_0^\infty \frac{dp}{1+\exp[(\varepsilon_{mp}+g\sigma\mu_BB/2-\mu)/T]}. \tag{48}$$

A comparison of Eqs. (4) and (47) shows that this term only gives an oscillating component of the magnetic response with Fourier coefficients which depend on the flux. Calculations similar to those made in Sec. 2 show that $M_\sigma^{(2)}$ contributes to the monotonic and to the oscillating components of the response, where the monotonic component depends nonlinearly on the flux Φ.

To sum up, we can affirm that in all the nanostructures studied allowance for the spin–magnetic interaction destroys the periodicity of the magnetic response as a function of the flux because the Fourier coefficients depend on the field B and a monotonic component appears in the response.

Since in real structures we find $m_0/m^*\gg 1$, the spin–magnetic interaction will be important in nanostructures for which the carrier has a large g-factor. To conclude, we note that for a bracelet the graphs depend strongly on the ratio of ρ to L (Fig. 3). The cylindrical nanostructures considered above generally exhibit oscillations of the magnetic response as a function of the flux. These are Aharonov–Bohm oscillations, i.e., they are periodic with respect to the magnetic field flux if the electron spin is neglected. We also note that for carbon nanotubes which do not possess complete cylindrical symmetry, flux periodicity also occurs.[2,19] In view of this observation, the cylinder model analyzed in Sec. 2 is clearly a reasonable approximation to describe the magnetic response of carbon nanotubes neglecting spin. If the electron spin is taken into account, the cylinder model does not give a crescent-shaped singularity on the $M(\Phi)$ curves, in contrast to the results obtained in Ref. 2.

This work was supported by Grants from the Russian Fund for Fundamental Research, the Russian Ministry for General and Professional Education, and the ''Russian Universities—Fundamental Research'' Program.

[1] H. Ehrenreich and D. Turnbull (Eds.), *Solid State Physics: Semiconductor Heterostructures and Nanostructures*, Vol. 44 (Academic Press, New York, 1991).
[2] M. F. Lin and K. W.-K. Shung, Phys. Rev. B **52**, 8423 (1995).
[3] H.-F. Cheung, Y. Gefen, E. K. Riedel, and W.-H. Shin, Phys. Rev. B **37**, 6050 (1988).
[4] Y. Meir, O. Entin-Wohlman, and Y. Gefen, Phys. Rev. B **42**, 8531 (1990).
[5] R. Merlin, Solid State Commun. **64**, 99 (1987).
[6] V. A. Geĭler, V. A. Margulis, and I. V. Chudaev, Zh. Éksp. Teor. Fiz. **109**, 762 (1996) [JETP **82**, 409 (1996)].
[7] V. A. Geĭler, V. A. Margulis, and O. B. Tomilin, JETP Lett. **63**, 578 (1996).
[8] V. A. Geyler and V. A. Margulis, Phys. Rev. B **55**, 2543 (1997).
[9] M. Buttiker, Y. Imry, and R. Landauer, Phys. Lett. A **96**, 365 (1983).
[10] M. Buttiker, Phys. Rev. B **32**, 1846 (1985).
[11] C. L. Foden, M. L. Leadbeater, J. H. Burroughes, and M. Pepper, J. Phys.: Condens. Matter **6**, L127 (1994).
[12] C. L. Foden, M. L. Leadbeater, and M. Pepper, Phys. Rev. B **52**, 8646 (1995).
[13] H. Aoki and H. Suezawa, Phys. Rev. A **46**, R1163 (1992).
[14] V. M. Nabutovskiĭ and D. A. Romanov, Zh. Éksp. Teor. Fiz. **90**, 232 (1986) [Sov. Phys. JETP **63**, 133 (1986)].
[15] V. V. Rotkin and R. A. Suris, Fiz. Tverd. Tela. (St. Petersburg) **36**, 3569 (1994) [Phys. Solid State **36**, 1899 (1994)].
[16] L. I. Magarill, D. A. Romanov, and A. V. Chaplik, JETP Lett. **64**, 460 (1996).
[17] L. I. Magarill, D. A. Romanov, and A. V. Chaplik, Zh. Éksp. Teor. Fiz. **113**, 1411 (1998) [JETP **86**, 771 (1998)].
[18] Yu. N. Ovchinnikov, W. Lehle, and A. Schmid, Ann. Phys. **6**, 489 (1997).
[19] H. Ajiki and T. Ando, J. Phys. Soc. Jpn. **62**, 1255 (1993).
[20] D. Shoenberg, *Magnetic Oscillations in Metals* (Cambridge University Press, Cambridge, 1984) [Russ. transl., Mir, Moscow, 1986].
[21] L. D Landau and E. M. Lifshitz, *Statistical Physics*, Part 1, 3rd ed., Pergamon, Oxford (1980) [Russian orig., Nauka, Moscow (1976)].
[22] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series*, Vols. 1–3 (Gordon and Breach, New York, 1986, 1986, 1989) [Russ. original, Vols. 1–3, Nauka, Moscow, 1981, 1983, 1986].
[23] *Higher Transcendental Functions (Bateman Manuscript Project)*, Vol. 1, edited by A. Erdélyi (McGraw-Hill, New York, 1953) [Russ. transl., later ed., Nauka, Moscow, 1965].
[24] A. I. Anselm, *Introduction to Semiconductor Theory* (Prentice Hall, Englewood Cliffs, N.J., 1981) [Russ. transl., Nauka, Moscow, 1978].

Translated by R. M. Durham

Statistical theory of a solvated electron in an electrolyte

G. N. Chuev*)

Institute of Mathematical Biology, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

(Submitted 10 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1463–1477 (April 1999)

The behavior of a solvated electron in an electrolyte is investigated. The formalism of the theory is based on variational estimation of path integrals. It reduces the problem to the investigation of the self-consistent mean field produced by the ions and the electron. Mayer cluster expansions make it possible to take account of the short-range interactions and to find expressions for the effective potential of the electron and the electron–ion and electron–neutral atom correlation functions as a function of the macro- and microscopic parameters of electrolytes. In the limit of high ion densities the behavior of the electron is determined solely by the Coulomb interaction, which results in the formation of a polaron state. This state of the electron is virtually independent of the thermodynamic parameters of the electrolyte. In the opposite limit of low ion densities the electron forms a cavity state. The presence of ions results in additional localization of the electron and is manifested experimentally as a shift of the absorption band in the direction of high energies. The estimated shift for a hydrated electron agrees with the experimental data. © *1999 American Institute of Physics.* [S1063-7761(99)02104-6]

1. INTRODUCTION

Solvated electrons—excess electron in liquid or gaseous media that do not form chemical bonds—are objects of both intense theoretical investigations and numerical simulation (see the reviews Refs. 1–5). Such a mixed quantum–classical system is convenient for demonstrating the possibilities of various numerical methods (quantum molecular dynamics, path integrals, various combined schemes). In the last few years there have appeared a large number of works[6–13] (see also the reviews Refs. 14 and 15) where the solvated-electron problem is studied on the basis of a statistical theory.

Experimental methods have now been developed and a great deal of experimental data on the behavior of a solvated electron in various media have been accumulated.[16–21] The behavior of an electron in liquid electrolytes stands out especially among the diverse experimental facts. In liquid electrolytes two kinds of interactions compete: the long-range Coulomb attraction and the short-range repulsion between the electron and the particles of the liquid. This competition causes a solvated electron in an electrolyte to depend strongly on the ion density and is fundamentally different for strong and weak electrolytes.

The present paper is devoted to the statistical theory of a solvated electron in an electrolyte. The formalism of this theory, based on the method of path integrals, makes it possible to reduce the problem to the investigation of the self-consistent mean field produced by the electron, after which the statistical approaches developed in the theory of liquids can be applied. Variational estimations of the path integrals makes it possible to determine the physics of the behavior of the electron at the microscopic level and to obtain ''almost analytically'' how the structural and energy characteristics of a solvated electron depend on the macro- and microscopic parameters of the liquid (density, temperature, pressure, size and charge of molecules, and so on).

The formalism of the method is presented in Sec. 2. In this method the problem is reduced to the investigation of the partition function of the grand canonical ensemble of the system. This partition function is given in terms of a path integral over the electric field induced by the charges in the liquid. In Sec. 3 this integral is evaluated in the mean-field approximation, which makes it possible to find the effective potential for a solvated electron and the electron–ion and electron–neutral atom binary functions as a function of the state of the electrolyte. Two limiting cases—low and high ion densities in the electrolyte—are studied in Secs. 4 and 5, respectively. Various approximations for the binary correlation functions, such as the random-phase approximation, the hyperchain approximation, or the Percus–Yevick approximation, make it possible to find the behavior of the electron in these two limiting cases and to establish a relation between the state of the electron and the thermodynamic parameters of the electrolyte. Section 6 is devoted to a discussion of the results obtained. A derivation of the expression for the partition function of the grand canonical ensemble in terms of a path integral over the electric field is presented in the Appendix.

2. FORMALISM OF THE METHOD

Let us consider a solvated electron in a classical liquid. The atoms of the liquid with which this electron interacts create a complicated potential field for the electron. The detailed analysis of this field is an extremely difficult problem. However, the existence of a large parameter $\tilde{N}$—the number

of interacting atoms of the liquid—makes it possible to assume this potential field to be random, to perform self-averaging of this field, and to find the behavior of the solvated electron as a function of the average parameters of the liquid. In this statistical approach the problem reduces to calculating the partition function of the grand canonical ensemble. For a solvated electron in a classical liquid the partition function Ξ of the grand canonical ensemble can be expressed in terms of a configurational integral, which depends on the configuration of the classical particles $\mathbf{R}_1,\mathbf{R}_2,\ldots=\mathbf{R}^{\{N\}}$, and a path integral over the electron coordinates $\mathbf{r}(\tau)$:

$$\Xi=\sum_{N\geqslant 0}^{\infty}\frac{z^N}{N!}Q_{N+1},\quad z=\left(\frac{2\pi M}{\beta}\right)^{-3/2}\exp(\beta\mu)V,$$

$$Q_{N+1}=\int D[\mathbf{r}(\tau)]\int d\mathbf{R}^{\{N\}}\exp\left\{-BU_{ss}-\int_0^\beta d\tau\left[\frac{1}{2}\mathbf{r}^2(\tau)+\sum_i^N u(\mathbf{r}(\tau)-\mathbf{R}_i)\right]\right\}. \tag{1}$$

Here $U_{ss}(\mathbf{R}^{\{N\}})$ is the interaction potential between the particles of the liquid, $u(\mathbf{r}-\mathbf{R}_i)$ is the pair interaction potential between an electron and a particle of the liquid, and $\mathbf{R}_i$ are the coordinates of the i-th classical particle. In the relation presented $k_BT=1/\beta$ is the temperature (we employ the system of units with $\hbar=1$, $m=1$, and $e=1$), V is the volume of the system, and μ and M are the chemical potential and the mass of the particles.

The problem of determining the state of a solvated electron reduces to calculating the integral (1). The dimension of this integral is very large. Therefore the main problem of the theoretical analysis is to decrease this dimension in a reasonable manner while preserving all interesting physical properties of the system under study.

We shall assume that the liquid in which the electron is solvated is an electrolyte with density ρ. This electrolyte contains both particles with charge ± 1, whose relative number is c, and neutral atoms, whose density and coordinates are $(1-c)\rho$ and $\mathbf{r}_{i0}$, respectively. The interaction potential U_{ss} includes a short-range repulsive part U_r of the hard-sphere-potential type and a long-range Coulomb interaction $u_{\pm q}=\pm u_q=\pm 1/|\mathbf{R}_{iq}-\mathbf{R}_{jq}|$ for charged particles with the coordinates $\mathbf{R}_{iq}$ and $\mathbf{R}_{jq}$ and charges ± 1. To simplify the calculations we assume the permittivity of the solvent to be $\varepsilon_0=1$. The influence of a permittivity $\varepsilon_0\neq 1$ can be taken into account by renormalizing the charge as $q^2\to q^2\varepsilon_0^{-1}$.

In the present paper we shall confine our attention to only two types of electron–atom interactions: a short-range repulsive potential u_0 between the electron and a neutral particle and a Coulomb interaction $u_{e\pm}$ between the electron and a charged particle:

$$u_0(r\leqslant d)=V_0>0,\quad u_0(r>d)=0,$$
$$u_{e\pm}(\mathbf{r}-\mathbf{R}_{iq})=\pm|\mathbf{r}-\mathbf{R}_{iq}|^{-1}. \tag{2}$$

We shall characterize the state of the solvated electron by a wave function $\phi(\mathbf{r})$. For simple estimates we shall employ a Gaussian wave function

$$\phi(r)=(2\alpha/\pi)^{3/2}\exp[-\alpha^2r^2]. \tag{3}$$

We note that for a solvated electron $d\alpha\ll 1$ because of the short range of the interaction.

Note that we have $\tilde{N}^{-1}=\rho\alpha^{-3}\ll 1$, since the solvated electron interacts with a large number of particles of the liquid. In the opposite case the state of the electron must be calculated not by statistical but by quantum-chemical methods.

The first step in decreasing the dimension of the integral (1) is to switch from a path integral over the electron coordinates $\mathbf{r}(\tau)$ to averaging over the electron distribution. The result is

$$Q_{N+1}=\int d\mathbf{R}^{\{N\}}\exp\left\{-\beta U_{ss}(\mathbf{R}^{\{N\}})-\beta T_e-\beta\sum_i^{cN}U_{e\pm}(\mathbf{R}_{iq})-\beta\sum_i^{(1-c)N}U_0(\mathbf{R}_{i0})\right\}, \tag{4}$$

where T_e is the kinetic energy of the electron and $U_{e\pm}$ and U_0 are defined as

$$U_{e\pm}(\mathbf{R}_{iq})=\pm\int\phi^2(\mathbf{r})|\mathbf{r}-\mathbf{R}_{iq}|^{-1}d\mathbf{r},$$
$$U_0(\mathbf{R}_{i0})=\int\phi^2(\mathbf{r})u_0(|\mathbf{r}-\mathbf{R}_{i0}|)d\mathbf{r}\simeq 4\pi V_0\phi^2(R_{i0})d^3/3, \tag{5}$$

and the choice of sign in $U_{e\pm}$ depends on the ion charge. The last two terms in Eq. (4) can be regarded as the long-range part ($U_e=(U_{e+}-U_{e-})/2$) and the short-range part (U_0) of an external field acting on the electrolyte. Thus, the problem reduces to finding the partition function for an electrolyte in an external field U_e+U_0.

To calculate Ξ we perform a transformation (see Appendix) and express the partition function of the grand canonical ensemble in terms of a path integral over the electric field Ψ:

$$\Xi=\Xi_0\int D[\Psi]\exp[-\beta\Omega\{\Psi,\phi\}], \tag{6}$$

$$\Omega=T_e+\frac{1}{2}(\Psi-U_e)*u_q^{-1}*(\Psi-U_e)-\beta^{-1}\delta N(\Psi,U_0), \tag{7}$$

$$\delta N(\Psi,U_0)=c\rho*f_q+\frac{c^2\rho^2}{2!}f_q*h_s*f_q+(1-c)\rho*f_s+\frac{(1-c)^2\rho^2}{2!}f_s*h_s*f_s+\frac{2c(1-c)\rho^2}{2!}f_q*h_s*f_s. \tag{8}$$

In these relations Ξ_0 is a normalization constant, Ω is the thermodynamic potential, $u_q^{-1}=-\Delta(\mathbf{r})$ is the inverse of the operator $u_q(\mathbf{r})$, and the symbol $*$ denotes a convolution

$$y*x\equiv\int x(\mathbf{r}_1)y(\mathbf{r}-\mathbf{r}_1)d\mathbf{r}_1.$$

In the relation (8) $h_s(\mathbf{r})$ is the complete density–density correlation function for a hard-sphere liquid, and f_q and f_s are Mayer functions

$$f_q=\frac{1}{2}[\exp(e\beta\Psi)+\exp(-e\beta\Psi)-2],$$

$$f_s=\exp(-\beta U_0)-1. \tag{9}$$

In the relation (7) the quantity δN is the change in the distribution of the particles as a result of the presence of the external field. The second term in Eq. (7) can be rewritten as

$$(\Psi-U_e)*u_q^{-1}*(\Psi-U_e)$$
$$=\int(\nabla\Psi)^2d\mathbf{r}+\int\frac{\phi^2(\mathbf{r}_1)\phi^2(\mathbf{r})}{|\mathbf{r}-\mathbf{r}_1|}d\mathbf{r}_1+d\mathbf{r}$$
$$+\int\Psi(\mathbf{r})\phi^2(\mathbf{r})d\mathbf{r}. \tag{10}$$

In this form the physical meaning of these contributions is quite clear. The first term in Eq. (10) corresponds to the internal energy of the electric field produced by all charges in the electrolyte. The second term is the energy of the electron in the field produced by the electron itself. The last term in the energy of the excess charge with density $\phi^2(\mathbf{r})$ in the field $\Psi(\mathbf{r})$.

At first glance it appears that we have complicated the problem, having written the partition function of the grand canonical ensemble in the form of an infinite-dimensional path integral. However, in contrast to the path integral in Eq. (1), the integral (6) specifies the dependence of the partition function on the classical electric field Ψ. This field is essentially a collective variable. To calculate the expression (6) we can employ estimates of multidimensional integrals, specifically, the method of steepest descent.

3. THE MEAN-FIELD APPROXIMATION

The excess electron induces in the electrolyte a mean electric field $\tilde{\Psi}$, whose Fourier components are related to the external field U_e via the permittivity ε, i.e., $\tilde{\Psi}(\mathbf{k})=\epsilon^{-1}(\mathbf{k})U_e(\mathbf{k})$. Since Eq. (5) implies $U_e\propto\alpha$, the argument of the exponential in Eq. (6) contains the parameter $\beta\alpha$, which in our case is large, $\beta\alpha\gg1$. For this reason we shall estimate (6) by the method of steepest descent, which yields the mean field $\tilde{\Psi}$ so that

$$\left.\frac{\partial\Omega}{\partial\Psi}\right|_{\Psi=\tilde{\Psi}}=0. \tag{11}$$

Then we obtain from Eqs. (7) and (8) a nonlinear Poisson–Boltzmann differential equation

$$u_q^{-1}*(\tilde{\Psi}-U_e)=\frac{c\rho}{2}(\exp[-\beta\tilde{\Psi}]-\exp[\beta\tilde{\Psi}])$$
$$\times(1+c\rho h_s*f_q+(1-c)\rho h_s*f_s). \tag{12}$$

We shall find the binary correlation functions $g_{e+}(r)$, $g_{e-}(r)$, and $g_{e0}(r)$, which describe the probability of finding a corresponding ion or neutral atom at a distance r from the center of localization of the electron:

$$g_{e\pm}(r)=-\frac{\partial\ln Q_{N+1}}{\partial\beta U_{e\pm}}=\exp(\pm\beta\tilde{\Psi})$$
$$\times(1+\rho h_s*f_q+(1-c)\rho h_s*f_s),$$
$$g_{e0}=-\frac{\partial\ln Q_{N+1}}{\partial\beta U_0}=\exp(-\beta U_0)$$
$$\times[1+(1-c)\rho h_s*f_s+c\rho h_s*f_q]. \tag{13}$$

The relation (12) can be rewritten as

$$\tilde{\Psi}=U_e-\rho_q*u_q, \tag{14}$$

where we have introduced the charge density $\rho_q(\mathbf{r})=c\rho g_{eq}(\mathbf{r})/2=c\rho(g_{e+}(\mathbf{r})-g_{e-}(\mathbf{r}))/2$. The equation (14) determines the mean field $\tilde{\Psi}(r)$ in terms of the external field U_e, which depends on the electron density distribution $\phi^2(r)$ and the correlation functions $g_{e+}(\mathbf{r})$ and $g_{e-}(\mathbf{r})$, which are related to the mean field $\tilde{\Psi}(\mathbf{r})$ by the relation (13). The integral equation (14) is the analog of the Ornstein–Zernike relation. To solve it, additional closure of Eq. (13) is required. Various modifications of this closure exist in statistical physics: the Percus–Yevick approximation, the hyperchain approximation, and so on.

The condition for a minimum of the variation of the thermodynamic potential as a function of the electron density

$$\frac{\partial\Omega}{\partial\phi}=0 \tag{15}$$

yields a nonlinear Schrödinger equation for the electron wave function $\phi(r)$

$$\left[-\frac{1}{2}\Delta+V_{\text{eff}}(r,\{\phi\})+E\right]\phi(r)=0, \tag{16}$$

where E is the electron energy and $V_{\text{eff}}(r,\{\phi\})$ is the effective potential for the electron,

$$V_{\text{eff}}(r,\{\phi\})=-\rho_q*u_q+(1-c)\rho g_{e0}*u_0. \tag{17}$$

The relations (13)–(14) and (16)–(17) form a closed system of equations for finding the electron wave function $\phi(\mathbf{r})$ as a function of the thermodynamic and structural parameters of the electrolyte: density, temperature, charge density, size of the atoms, structure factor, and so on.

The general case can be investigated only numerically. We shall investigate two limiting cases: weak ($c\ll1$) and strong ($c\simeq1$) electrolytes.

4. STRONG ELECTROLYTE

In this case we have $c\simeq1$, and the effect of the uncharged particles of the liquid can be neglected. This state corresponds to melts of the salts KCl, NaCl, and so on. The subsequent investigation depends on the type of approximation for closure of Eq. (13).

4.1 Random-phase approximation

In this case the linear approximation

$$g_{e\pm}\simeq(1+\rho h_s)(1\pm\beta\tilde{\Psi}), \tag{18}$$

is used instead of Eq. (13). This yields the solution for the Fourier component $\tilde{\Psi}(k)$

$$\tilde{\Psi}(k)=\frac{4\pi\phi^2(k)}{k^2+\kappa_D^2[1+\rho h_s(k)]},$$

$$\rho_q(k)=\frac{[1+\rho h_s(k)]\phi^2(k)\kappa_D^2}{k^2+\kappa_D^2[1+\rho h_s(k)]}, \tag{19}$$

where $\phi^2(k)$ is a Fourier component of the electron wave function and $\kappa_D=(4\pi\rho\beta)^{1/2}$ is the reciprocal of the Debye radius. For the simplest estimates the approximation $1+\rho h_s(k)\simeq\cos(\sigma k)$, where σ is the size of the particles of the liquid, can be used. We also note that in most liquids the characteristic radius of a solvated electron is $\alpha^{-1}\sim\sigma\sim 3$ Å, while at temperatures $T\sim 300-1500$ K and $\rho\sigma^3\sim 0.9$ we have $\sigma\kappa_D=230-150\gg 1$, i.e., the strong-electrolyte approximation corresponds to the strong-screening approximation, for which

$$\tilde{\Psi}(k\to 0)\simeq 4\pi\phi^2(0)\kappa_D^{-2},\quad \rho_q(k\to 0)\simeq\phi^2(0).$$

Correspondingly, for $r\to 0$ we obtain

$$g_{e\pm}(r)\simeq 1\pm\phi^2(r)\rho^{-1},$$

$$V_{\text{eff}}(r)\simeq-\int\phi^2(\mathbf{r}_1)|\mathbf{r}-\mathbf{r}_1|^{-1}d\mathbf{r}_1.$$

The behavior of the mean field $\tilde{\Psi}(r)$ depends on the ratio of the size of the particles of the liquid and the solvated electron. For the case $\sigma\alpha\ll 1$

$$\tilde{\Psi}(r)\simeq 4\pi\phi^2(r)\kappa_D^{-2}.$$

For the more realistic situation where $\sigma\alpha\simeq 1$, $\tilde{\Psi}(r)$ can be estimated according to the theory of residues as

$$\tilde{\Psi}(r)\propto\sin(\pi r/2\sigma)(\kappa_D^2\sigma^2 r)^{-1}+\ldots,$$

i.e., the field is an oscillating function with period $\sim\sigma^{-1}$.

Using the Gaussian approximation (3) for the wave function $\phi(r)$, from the relation (15) we obtain a condition for α:

$$3\alpha a_0=\pi^{-1/2}, \tag{20}$$

where a_0 is the Bohr radius, which gives $\alpha\simeq 0.355$ Å^{-1} and $T_e=(3/2)\alpha^2=0.106$ a.u. It can be shown that in this limit the kinetic energy T_e, potential energy P_e, and total energy E of the electron satisfy the virial relation,

$$|T|:|E|:|\Pi_e|=1:3:4. \tag{21}$$

Using the relations presented above, we calculated the indicated energy characteristics (T_e,P_e,E) for a solvated electron in KCl melt at $T=1000$ K (see Table I). We note that the variational estimate of E differs by only 2% from the estimate $E=-0.324$ a.u. obtained by solving the Schrödinger equation numerically.[22] To estimate the absorption band maximum $E_{\max}=E_1-E$ we employed the data of Ref. 22, according to which for the Schrödinger equation (16) with the potential

$$V_{\text{eff}}(\mathbf{r})\simeq-\int\phi^2(\mathbf{r})|\mathbf{r}-\mathbf{r}_1|^{-1}d\mathbf{r}_1$$

TABLE I. Kinectic T_e, potential P_e, and total E energies of a solvated electron in KCl melt, together with the absorption band maximum $E_{\max}$.

	T_e, a.u.	P_e, a.u.	E, a.u.	$E_{\max}$, a.u.
Theory				
Random-phase approximation	0.106	−0.424	−0.318	0.131
RISM-polaron[7]	0.120	−0.450	−0.330	0.161[24]
Simulation[23]	0.09±0.026	−0.41±0.064	−0.314	0.125,[24] 0.132[25]
Experiment[26]				0.098

the energy of the excited state is $E_1=-0.187$ a.u. Table I also gives the characteristics of an electron obtained by calculating the integrals over the fields[23] numerically and in the RISM (reference interaction site model)–polaron model.[7] The absorption maximum was also estimated in Ref. 24 by analytic continuation for the generalized susceptibility using the results of the RISM–polaron model and data obtained from a direct calculation of the path integrals.[23] Table I also gives an estimate obtained for the absorption maximum by solving the time-dependent Schrödinger equation numerically.[25] As one can see from the table, our theory agrees very well with the numerical-simulation data both with respect to energy and the potential maximum. However, the experimental value of the absorption maximum[26] is somewhat lower than the values obtained from the theory and the simulation. We note that the calculations presented[7,23] employed a potential $u_{e+}(r)$ that is different from Eq. (2), specifically,

$$u_{e+}(r<R_c)=R_c^{-1},\quad u_{e+}(r\geqslant R_c)=r^{-1}.$$

The virial relation (21) for the energies has been observed in a direct calculation of the path integrals for an excess electron in KCl melt.[23] Relations (21) were satisfied to better than the computational error in the quantities themselves. The numerical deviations from the value $T_e\simeq(3/2)\alpha^2=0.106$ a.u. lie within the same limits and are related to the deviations of the electron–ion interaction u_{e+} as $r\to 0$ from the Coulomb law.

4.2 Hyperchain approximation

In this approximation Eq. (13) is modified to the form

$$g_{e\pm}\simeq\exp[\pm\beta\tilde{\Psi}]\simeq\exp[\pm\beta(U_e-\rho h_{ss}*\phi^2*c_e)]. \tag{22}$$

This relation follows from Eq. (14) if the approximation

$$\rho_q\simeq\rho h_{ss}*\phi^2,$$

is used, where

$$\rho h_{ss}(k)=-[1+\rho h_s(k)]\kappa_D^2/(k^2+\kappa_D^2[1+\rho h_s(k)])$$

is the total correlation function of the electrolyte. Then

$$\rho_q*u_q\simeq\rho h_{ss}*\phi^2*c_e,$$

where c_e is the direct electron–ion correlation function.

To determine $\tilde{\Psi}(r)$, in this case, Eqs. (14) and (22) must be solved numerically. This requires a separate investigation. Here we shall consider the case $h_s \equiv 0$. Then we find from Eq. (22)

$$g_{e\pm} \simeq \exp[\pm \varphi^2(r)\rho - 1]. \tag{23}$$

Other nonlinear corrections can be taken into account in this limit. Since $\tilde{N}^{-1} \leqslant 1$ holds, we obtain from Eq. (13)

$$g_{e\pm} \simeq \exp\left[\pm\left\{\varphi^2(r)\rho^{-1} - \frac{1}{3!}\varphi^6(r)\rho^{-3} + \ldots\right\}\right]. \tag{24}$$

The relations (23) and (24) only change the behavior of the correlation functions $g_{e\pm}(r \to 0)$. They do not affect the energy characteristics and the virial relation (21) or the average radius of the electron distribution.

We note one other important aspect. A quantitative measure expressing whether repulsion or attraction effects dominate is the change ΔN in the average number of molecules bound on the electron as compared with a uniform liquid:

$$\Delta N = \int [c\rho(g_{e+} + g_{e-} - 2) + (1-c)\rho(g_{e0} - 1)]d\mathbf{r}. \tag{25}$$

For the hyperchain approximation in the case of a strong electrolyte $\Delta N > 0$ holds, i.e., the attractive forces dominates. Then the state of the electron is similar to a polaron state,[27] except that instead of a phonon ''coat'' the electron is bound with ions forming a cluster.

The short-range repulsion between particles of the liquid, just like the short-range interaction between ions and the electron, which is not studied in the present paper, will influence the behavior of a solvated electron, but this influence will be small for the energy characteristics because of the short range of the interaction. The effect of the nonpolar particles or particles with a dipolar charge will also be weak if their density is low.

In summary, the behavior of a solvated electron in a strong electrolyte is universal and is determined by the Coulomb interaction. Its characteristic size is $\alpha^{-1} = (3\sqrt{\pi})^{-1} a_0 \simeq 3$ Å, and the energy characteristics satisfy the virial relation (21) and are virtually independent of the temperature, density, and other parameters of the liquid.

5. WEAK ELECTROLYTE

In this case the interaction of the electron with uncharged particles plays the main role. In the limit $c \to 0$ we obtain from Eq. (7)

$$\Omega = T_e - \rho * f_s - \frac{1}{2!} f_s * \rho^2 h_s * f_s. \tag{26}$$

Using the estimate (5) for u_0, we find the condition for the radius of the solvated electron

$$3\alpha \simeq \frac{4\pi C_0 \rho}{\beta \alpha^4}, \tag{27}$$

where $C_0 \sim 1$ is a numerical parameter. At room temperatures and $\rho\sigma^3 \simeq 0.9$ we have $\alpha\kappa_D^{-1} \gg 1$ for $c \gg 1$, i.e., the weak-electrolyte approximation corresponds to weak screening. In this case $\Delta N < 0$ holds, i.e., the repulsive forces predominate and the electron forms a cavity with characteristic radius α^{-1}.

The correlation functions g_{eq} and g_{e0} are found from the conditions

$$g_{eq} \simeq (\exp(\beta\tilde{\Psi}) - \exp(-\beta\tilde{\Psi}))[1 + \rho h_s * f_s],$$

$$g_{e0} \simeq \exp(-\beta U_0)[1 + \rho h_s * f_s],$$

$$\tilde{\Psi}(r) \simeq U_e + \ldots \simeq \min(\alpha, r^{-1}). \tag{28}$$

The estimate of the behavior of the solvated electron depends on the type of closure chosen. For the random-phase approximation, where the approximation (18) can be used, we obtain

$$\tilde{\Psi}(k) = \frac{4\pi\phi^2(k)}{k^2 + \kappa_r^2},$$

$$\kappa_r^2 = 4\pi\rho\beta(1 + h_s(k)f_s(k)) \simeq 4\pi\rho\beta[1 + h_s(0)f_s(0)], \tag{29}$$

$$g_{eq}(r \to 0) \simeq \frac{2\beta}{r}\exp[-\kappa_r r],$$

$$V_{\text{eff}}(r) \simeq \frac{4\pi}{3} V_0 \rho d^3 g_{e0}(r). \tag{30}$$

The corrections associated with the presence of charged particles will modify the radius of the solvated electron, which will be found from the relation

$$3a_0\alpha = \frac{4\pi C_0 \rho}{\beta\alpha^4} + \frac{c\kappa_r^2}{2\sqrt{\pi}\alpha^2} - \frac{4\pi C_0 c\rho}{\beta\alpha^4}. \tag{31}$$

It is easy to see that the last term is small compared with the second term, and their ratio is $\sim(\beta\alpha)^{-2} \ll 1$ for ordinary temperatures. Thus, the introduction of charged particles decreases the radius of a solvated electron, i.e., it results in additional localization and increases the kinetic energy of the electron. This should be observed experimentally as a shift of the absorption band in the short-wavelength direction.

In the hyperchain approximation $g_{e\pm} \simeq \exp(\pm\beta U_e)$, which gives $g_{e+}(r \to 0) \simeq \delta(r)$, i.e., localization of the electron in a positive field and appearance of a chemical bond occur. This result is a consequence of neglecting the short-range interaction forces between the electron and the ions. The existence of even a weak repulsive potential will cause smearing of the electronic density. Strictly speaking, the problem of an electron–ion complex should be studied quantum-chemically.

On the whole, the calculation of the absorption spectrum of an excess electron in a weak electrolyte requires taking into account in detail the contributions of short-range forces, Debye screening, electronic polarization, dipole interactions and so on. In the numerical simulation of a hydrated electron[28] a correlation was noted between the energy $E_{\max}$ of the absorption maximum and the hydration radius r_h of the electron. We approximate this correlation as $E_{\max} \propto r_h^{-1}$. Taking account of this fact and the relation (31) we find that

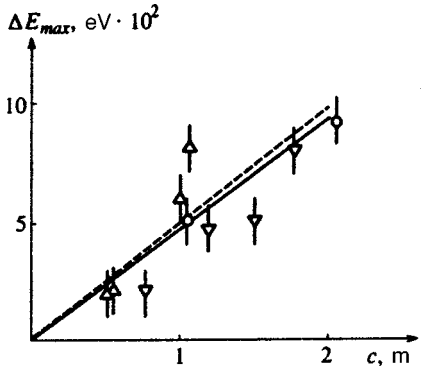

FIG. 1. Experimental data on the shift of the absorption maximum of a hydrated electron in aqueous solutions of LiCl (▽), $MgCl_2$ (○), and $LiClO_4$ (△) as functions of the solution concentration.[29] The solid line corresponds to a linear fit of the data and the dashed line is the theoretical estimate.

the relative shift $\Delta E_{\max}$ of the absorption maximum depends on ratio of the Debye radius and the hydration radius of the electron

$$\frac{\Delta E_{\max}}{E_{\max}}=\frac{2}{9\sqrt{\pi}}c\kappa_D^2 r_h^2=\frac{8\sqrt{\pi}}{9\epsilon_0}c\beta\rho r_h^2. \tag{32}$$

Here we took account of the fact that

$$r_h^2=\int r^2\phi^2(r)d\mathbf{r}=3/4\alpha^2$$

and the permittivity of the solvent is $\epsilon_0\neq 1$. Using for water the experimental value $E_{\max}=1.72$ eV and the hydration radius $r_h=2.05\pm 0.1$ Å obtained by numerical simulation,[28] we find that at room temperature

$$\frac{\Delta E_{\max}}{c}=(4.9\pm 0.5)\times 10^{-2}\ \text{eV/M}.$$

Figure 1 shows the experimental values (symbols) for the shift of the absorption maximum of a hydrated electron in aqueous solutions of LiCl, $MgCl_2$, and $LiClO_4$ as a function of the solution concentration[29] for $c\leqslant 2.1$ M. Fitting these data by a linear relation we obtain

$$\frac{\Delta E_{\max}}{c}=(4.6\pm 1.7)\times 10^{-2}\ \text{eV/M}$$

(in the figure this dependence is shown by the solid line), which is in very good agreement with the theoretical estimate (dashed line). Therefore at low ion densities the relative shift of the absorption maximum for a hydrated electron is determined completely by the Debye screening in accordance with Eq. (32). At higher concentrations we have $\alpha\kappa_D^{-1}\sim 1$ and additional terms in Eq. (31) must be taken into account. In experiments at concentrations $c\geqslant 2.1$ M (Ref. 29) this gives rise to a nonmonotonic dependence $\Delta E_{\max}(c)$, associated with structural rearrangements (a polar liquid is a concentrated electrolyte) around the electron.

6. DISCUSSION

Using a statistical approach we studied the behavior of a solvated electron in an electrolyte. In the case of a strong electrolyte, in the limit of high ion densities, this behavior is determined only by the Coulomb interaction, which causes clustering of positive ions on the electron and the formation of a polaron localized state. This state of the electron is virtually independent of temperature, ion density, and other parameters of the liquid. The variational estimates made of the electron energy at the absorption band maximum agree well with the data from numerical simulation of a solvated electron in KCl melt,[23] but there is a discrepancy between these data and experiment. Moreover, experimentally, the absorption band maximum is observed to vary in the series of alkaline-halide solutions: It decreases from 2.2 eV (LiCl) to 1.07 eV (CsCl).[26] This fact can hardly be explained using the simple model considered here. In our view, the discrepancy is due primarily to the approximation used to calculate the effects associated with the electronic polarization of the medium and to deviation of the electron–ion interaction potential from a Coulomb potential.

In the opposite limit of low electron concentrations a cavity state is formed, where the characteristic size of the cavity is determined by the parameter $\alpha\propto\rho^{1/5}\beta^{-1/5}$. This behavior is typical for an electron in disordered systems with a short-range potential.[30] The introduction of a small number of charged particles causes additional electron localization and should be manifested experimentally as a shift of the absorption band in the direction of high energies. This shift of the absorption band was investigated theoretically in Refs. 31 and 32, and it has been observed experimentally in water[33] with increasing ion concentration. Our estimate of the relative shift of the absorption maximum for a hydrated electron agrees well quantitatively with these experimental data.

The mathematical basis of the method developed is conversion of the partition function of the grand canonical ensemble into a path integral over the electric field induced by the charges and the electrolyte. Such a transformation was probably first performed in Ref. 34. It was used in Ref. 35 to find the thermodynamic and structural parameters of a classical system in a long-range field. The Mayer cluster expansion method makes it possible to include a short-range interaction in the analysis and to obtain relations for the effective potential and the free energy of the solvated electron and the electron–ion and electron–neutral atom binary functions. The formalism for separating the short- and long-range interactions in the transformation performed has been investigated in detail in Ref. 36 by a diagrammatic technique.

In Sec. 2 we switched immediately to a description of the electron in terms of the electron density distribution function, after which we performed the indicated transformation. Strictly speaking, we should have proceeded the other way around: performing the transformation first and then estimating the path integral over the electron coordinates. If we had limited ourselves only to the quadratic term in the field, i.e., if we had employed the random functions approximation, then the path integral over the field could have been calculated analytically and the chemical potential of the electron could have been found explicitly. This method served as the basis for the RISM-polaron theory.[6] This theory has been used to calculate the behavior of a solvated electron in KCl

melt.[7] The latter means that on averaging we actually consider only the diagonal elements of the density matrix $\varrho(r,r) \propto \phi^2(r)$ and neglect the off-diagonal elements. For this reason, the order of the transformations is of no significance. If the short-range attraction predominates ($u_0 < 0$), it is important to take into account the off-diagonal disorder.

In our view, the proposed statistical method is a powerful tool for calculating at the microscopic level the structural and thermodynamic parameters of a solvated quantum particle in a classical liquid. It makes it possible to find the behavior of the quantum particle self-consistently as a function of the molecular structure and the thermodynamic state of the medium.

APPENDIX

Transformation of the partition function of the grand canonical ensemble

To simplify the calculations we shall assume that the masses and chemical potentials of the atoms and ions are identical. We introduce the generalize d charge density $\rho_q(r)$ and the neutral-atom density $\rho_0(r)$:

$$\rho_q(r) = \sum_i^{cN} \delta(r - R_{iq}), \quad \rho_0(r) = \sum_i^{(1-c)N} \delta(r - R_{i0}). \tag{A1}$$

Then the partition function Ξ of the grand canonical ensemble can be represented as

$$\Xi = \sum_{N>0} \frac{z^N}{N!} \int d\mathbf{R}^{\{N\}} \exp\left[-\beta(T_e + U_e) * \rho_q + \rho_q * \frac{u_q}{2} * \rho_q + U_0 * \rho_0 + U_s \right]. \tag{A2}$$

In this relation

$$d\mathbf{R}^{\{N\}} \equiv d\mathbf{R}_{iq}^{\{cN\}} d\mathbf{R}_{i0}^{\{(1-c)N\}},$$

and the short-range potential U_s depends on both the configuration $R_{i0}^{\{cN\}}$ of the ions and the configuration $R_{i0}^{\{(1-c)N\}}$ of the neutral particles. For the potential $A(r)$ belonging to the class of functions L_2, we can perform a Fourier transform for the exponential of a quadratic form:[37]

$$\exp\left[\frac{1}{2} \rho_q * A * \rho_q \right] = \left\{ \int D[\Psi] \times \exp\left[-\frac{1}{2} \Psi * A^{-1} * \Psi \right] \right\}^{-1} \int D[\Psi] \times \exp\left[-\frac{1}{2} \Psi * A^{-1} * \Psi + \rho_q * \Psi \right]. \tag{A3}$$

The Coulomb potential $u_q(r)$ belongs to this class of functions, and for it the inverse operator $u_q^{-1}(r) \equiv -\Delta(r)$ exists. Then we obtain from Eq. (A2)

$$\Xi = \Xi_0 \int D[\Psi] \exp\left[-\beta T_e - \frac{1}{2} \beta(\Psi - U_e) * u_q^{-1} * \times (\Psi - U_e) \right] I(\Psi, U_0),$$

$$I(\Psi, U_0) = \sum_{N \geqslant 0} \frac{z^N}{N!} d\mathbf{R}^{\{N\}} \exp\left[-\sum_i^{cN} \pm \beta\Psi(R_{iq}) - \sum_i^{(1-c)N} \beta U_0(R_{i0}) - \beta U_s \right]. \tag{A4}$$

The latter relation becomes

$$I = \sum_{N>0} \int d\mathbf{R}^{\{N\}} \frac{z^N}{N!} \prod_k^{cN} \exp[\pm \beta \Psi_{kq}] \times \prod_m^{(1-c)N} \exp[-\beta U_{m0}] \exp[-\beta U_s]. \tag{A5}$$

We now introduce the n-particle correlation functions of the electrolyte $\rho_q^{(n)}(r_1, \ldots, r_n)$ which characterize the probability of finding ions at the points $r_1, \ldots, r_n$ for a liquid consisting of hard spheres:[38]

$$\rho_q^{(n)}(r_1, \ldots, r_n) \propto \sum_{N>0} \frac{z^{N-n}}{(N-n)!} \int \exp[-\beta U_s] dR^{\{N-n\}}. \tag{A6}$$

Similarly we can find the correlation functions $\rho_0^{(n)}(r_1, \ldots, r_n)$ for the neutral atoms, and so on. Using the Mayer cluster expansions the relation (A5) can be put into the form

$$I = 1 + f_q * \rho_q^{(1)} + f_s * \rho_0^{(1)} + \frac{1}{2!} f_q * \rho_q^{(2)} * f_q + \frac{1}{2!} f_s * \rho_0^{(2)} * f_s + \frac{2}{2!} f_s * \rho_{0q}^{(2)} * f_q + \ldots + \frac{1}{n!} f_q * \rho_q^{(n)} * f_q \ldots * f_q + \ldots . \tag{A7}$$

We take account of the fact that

$$\rho_q^{(1)} = c\rho, \quad \rho_0^{(1)} = (1-c)\rho,$$
$$\rho_q^{(2)}(r) = c^2\rho^2[1 + h_s(r)],$$
$$\rho_0^{(2)}(r) = (1-c)^2\rho^2[1 + h_s(r)],$$
$$\rho_{0q}^{(2)}(r) = (1-c)c\rho^2[1 + h_s(r)],$$

where $h_s(r)$ is the total distribution function for a hard-sphere liquid. The relation (A7) can be written in the exponential form

$$I = \exp\left[cp * f_q + (1-c)\rho * f_s + \frac{c^2\rho^2}{2!} f_q * h_s * f_q + \frac{(1-c)^2\rho^2}{2!} f_s * h_s * f_s + \frac{2(1-c)\rho^2}{2!} f_s * h_s * f_q + a(h_s^{(3)}, \ldots) \right]. \tag{A8}$$

In this relation the term $a(h_s^{(3)}, \ldots)$ includes convolutions of third- and higher-order irreducible correlation functions for a hard-sphere liquid. Neglecting these irreducible correlations, we obtain Eq. (7).

This work was supported in part by the Russian Fund for Fundamental Research (Grant No. 98-01-01154).

*)E-mail: chu@impb.serpukhov.su

[1] J. Jortner and N. R. Kestner, *Electrons in Fluids* (Springer, Berlin, 1973).
[2] L. Kevan and B. Webster, *Electron-Solvent and Anion-Solvent Interactions* (Elsevier, Amsterdam, 1976).
[3] L. Kevan and D. F. Feng, Chem. Rev. **80**, 1 (1980).
[4] E. M. Itskovich, A. M. Kuznetsov, and J. Ulstrup, in *The Chemical Physics of Solvation*, edited by R. R. Dogonadze, E. M. Itskovich, E. Kalman, A. A. Kornyshev, and J. Ulstrup (Elsevier, Amsterdam, 1988).
[5] C. Ferradini and J. P. Jay-Gerin, *Excess Electrons in Dielectric Media* (CRC Press, Boca Raton, 1991).
[6] D. Chandler, Y. Singh, and D. M. Richardson, J. Chem. Phys. **81**, 1975 (1984).
[7] G. Malescio and M. Parrinello, Phys. Rev. A **35**, 897 (1987).
[8] M. R. Shaw and D. Thirumalai, J. Chem. Phys. **93**, 3450 (1990).
[9] J. Zhu and R. Cukier, J. Chem. Phys. **99**, 5384 (1993).
[10] S. Miura and F. Hirata, J. Chem. Phys. **98**, 9649 (1994).
[11] G. N. Chuev, Zh. Éksp. Teor. Fiz. **105**, 626 (1994) [JETP **78**, 334 (1994)].
[12] G. N. Chuev, Physica D **83**, 68 (1995).
[13] G. N. Chuev, Izv. Akad. Nauk, Ser. Fiz. **61(9)**, 1770 (1997).
[14] D. Chandler and K. Leung, Annu. Rev. Phys. Chem. **45**, 1500 (1994).
[15] G. N. Chuev, Usp. Fiz. Nauk **169**(1) (1999).
[16] E. J. Hart and M. Anbar, *The Hydrated Electron* (Wiley-Interscience, New York, 1970; Russian translation, Atomizdat, Moscow, 1973).
[17] J. Thompson, *Electrons in Liquid Ammonia* (Clarendon Press, Oxford, 1976; Russian translation, Mir, Moscow, 1979).
[18] A. K. Pikaev, *Modern Radiation Chemisty. Radiolysis of Gases and Liquids* (Nauka, Moscow, 1986).
[19] A. G. Khrapak and I. T. Yakubov, *Electrons in Dense Gases and Plasma* (Nauka, Moscow, 1981).
[20] J. H. Baxendale and F. Busi, *The Study of Fast and Transient Species by Electron Pulse Radiolysis* (Reidel, Dordrecht, 1982).
[21] R. M. A. Farhataziz, *Radiation Chemistry: Principles and Applications* (VCH, New York, 1987).
[22] N. K. Balabaev and V. D. Lakhno, Preprint, Scientific and Technial Information Branch of the Science Center for Biological Studies (Pushchino, 1979).
[23] M. Parrinello and A. Rahman, J. Chem. Phys. **80**, 860 (1984).
[24] G. Malescio, Phys. Rev. A **36**, 5847 (1987).
[25] A. Selloni, R. Carnevali, R. Car, and M. Darrinello, Phys. Rev. Lett. **59**, 833 (1987).
[26] J. F. Rounsaville and J. J. Lagowski, J. Phys. Chem. **72**, 1111 (1968).
[27] V. D. Lakhno and G. N. Chuev, Usp. Fiz. Nauk **165**, 285 (1995).
[28] P. J. Rossky and J. Schnitker, J. Phys. Chem. **92**, 4277 (1988).
[29] I. V. Kreĭtus and D. A. Strode, Khim. Vys. Énerg. **23**, 218 (1989).
[30] I. M. Lifshitz, Usp. Fiz. Nauk **83**, 617 (1964) [Sov. Phys. Usp. **7(4)**, 549 (1965)].
[31] V. D. Lakhno and O. V. Vasil'ev, Chem. Phys. Lett. **177**, 147 (1991).
[32] V. D. Lakhno and O. V. Vasil'ev, Phys. Lett. A **152**, 300 (1991).
[33] I. V. Kreitus, J. Phys. Chem. **89**, 1987 (1985).
[34] D. N. Zubarev, Dokl. Akad. Nauk SSSR **95**(4), 757 (1954).
[35] A. L. Kholodenko, J. Chem. Phys. **91**, 4849 (1989).
[36] I. R. Yukhnovskiĭ and M. F. Golovko, *The Statistical Theory of Classical Equilibrium Systems* (Naukova dumka, Kiev, 1980).
[37] M. V. Fedoryuk, *The Method of Steepest Descent* (Nauka, Moscow, 1977).
[38] T. Hill, *Statisical Mechanics: Principles and Selected Applications* (McGraw-Hill, New York, 1956; Russian translation, Izd. Inostr. Lit., Moscow, 1960).

Translated by M. E. Alferieff

Influence of spin–orbit interaction of two-dimensional electrons on the magnetization of nanotubes

L. I. Magarill and A. V. Chaplik*)

Institute of Semiconductor Physics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk
(Submitted 28 October 1998)
Zh. Éksp. Teor. Fiz. **115**, 1478–1483 (April 1999)

Calculations are made of the magnetization of a nanotube in a longitudinal magnetic field. It is shown that the spin–orbit interaction of two-dimensional electrons located at the surface of the nanotube causes a qualitative change in the magnetization. Depending on the parameters of the system, either diamagnetism or paramagnetism can occur and the dynamic susceptibility is characterized by anomalous dispersion at low frequencies. © *1999 American Institute of Physics.* [S1063-7761(99)02204-0]

1. INTRODUCTION

A two-dimensional electron gas at the surface of a circular cylinder is encountered in various experimental scenarios. For instance, the Aharanov–Bohm effect[1] under conditions of weak localization[2] is associated with such a gas. The magnetic properties of a thin-walled metal cylinder (in particular, superconducting) were investigated by Kulik[3] back in 1970. On the mesoscopic level particular mention should be made of carbon nanotubes in which two-dimensional conductivity can also occur. Most recently, Prinz *et al.*[4] developed an original method of ''folding'' strained GaAs/InAs layers to form cylinders and rolls with a radius of curvature of order a few tens or hundreds of angstrom.

In the present paper we investigate the magnetic properties of nanotubes in a field parallel to the axis of the cylinder. It is found that the spin–orbit interaction of $2D$ electrons at the surface of the cylinder is responsible for the qualitative behavior of the magnetization in static and variable external magnetic fields, even at very low frequencies. This last factor is important for experiments which use a modulation method to measure the magnetic susceptibility. The physical reason for these characteristics is associated with the crossing (or quasicrossing) of single-electron terms as a function of the magnetic flux.

2. MAGNETIC MOMENT OF A NANOTUBE IN A STATIC FIELD

We shall allow for the spin–orbit interaction using the Rashba model.[5] The corresponding Hamiltonian for a planar $2D$ system is written as

$$\hat{V}_{so}=\alpha[\hat{\sigma},\hat{\mathbf{p}}]\mathbf{n}, \tag{1}$$

where $\hat{\sigma}_i$ and $\hat{\mathbf{p}}$ are the Pauli matrices and the two-dimensional momentum operator, respectively, $\mathbf{n}$ is the normal to the surface, and α is the effective spin–orbit interaction constant. The constant α does not go to zero for an oriented surface on which two directions $\mathbf{n}$ are nonequivalent. It follows from Eq. (1) that in this system (electron gas at the surface of a circular cylinder in a magnetic field directed along the cylinder axis) the Hamiltonian has the form (we assume that the cylinder radius R is much greater than the lattice constant and we use cylindrical coordinates with the z axis directed along the axis of the cylinder)

$$\hat{H}_0=\frac{\hat{p}_z^2+(\hat{p}_\varphi+\Phi/R)^2}{2m}+\alpha[\hat{\sigma}_z(\hat{p}_\varphi)+\Phi/R)-\hat{\Sigma}\hat{p}_z)], \tag{2}$$

$$\hat{\Sigma}=\begin{bmatrix} 0 & ie^{-i\varphi} \\ -ie^{i\varphi} & 0 \end{bmatrix}. \tag{3}$$

Here p_z is the longitudinal momentum operator, $\hat{p}_\varphi=-i(1/R)\partial/\partial\varphi$, Φ is the magnetic flux across the cross section of the cylinder in units of the flux quantum $\Phi_0=2\pi\hbar c/e$, and e is the absolute electron charge; we shall take $\hbar$ equal to unity.

The Schrödinger equation with the Hamiltonian (2) can be solved exactly. The energy spectrum is given by

$$E_{j,\mu}(k)=B\left[k^2+\lambda_j^2+\frac{1-2\Lambda}{4}+\mu D_j\right], \tag{4}$$

where $D_j=\sqrt{\lambda_j^2(\Lambda-1)^2+k^2\Lambda^2}$, $B=1/2mR^2$, $k=p_zR$, j is the projection of the total momentum on the cylinder axis (half-integer), $\lambda_j=j+\Phi$, $\Lambda=2m\alpha R$, and $\mu=\pm1$ is a quantum number which labels the two branches of the spin-split dispersion law of each subband j. The normalized wave functions have the form (L is the cylinder length)

$$\Psi_{(\varphi)}=\frac{\exp(ip_zz)}{\sqrt{L}}\begin{pmatrix} \exp\{i(j-1/2)\varphi\} & \psi^{(1)} \\ \exp\{i(j+1/2)\varphi\} & \psi^{(2)} \end{pmatrix}, \tag{5}$$

where

$$\psi_{j^+}^{(1)}=\psi_{j^-}^{(2)}=iA_jC_j/\Lambda k, \quad \psi_{j^-}^{(1)}=\psi_{j^+}^{(2)}=A_j,$$

$$A_j=|\Lambda k|/\sqrt{4\pi D_jC_j}, \quad C_j=D_j+\lambda_j(\Lambda-1), \tag{6}$$

The spectrum and wave functions for the case $\Phi=0$ were obtained previously by Magarill *et al.*[6]

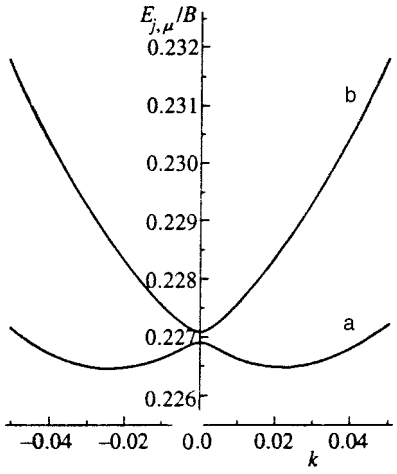

FIG. 1. Energy versus longitudinal momentum near half-integer value of the flux: $\Phi=0.4995$, $\Lambda=0.046$ (GaAs for $R=100$ Å). Curve a: $j=-1/2$, curve b: $j=-1/2$, $\mu=+1$. The gap disappears at $\Phi=1/2$.

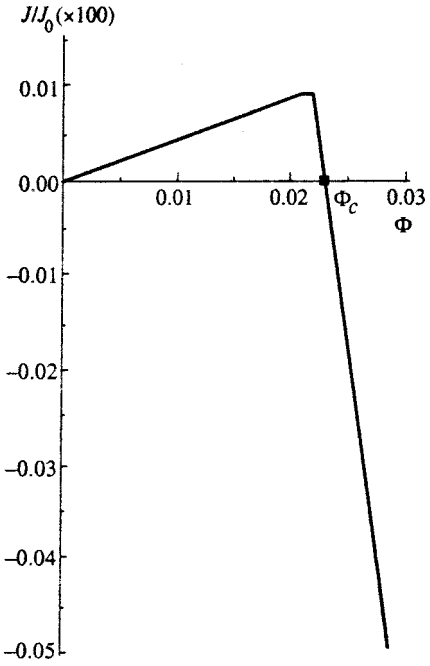

FIG. 2. Behavior of the equilibrium current for weak magnetic fluxes: $J_0=eB/\pi R$, $N_s=2.3\times10^9$ cm^{-2}, and the other parameters are the same as in Fig. 1.

The magnetic moment is proportional to the equilibrium current (persistent current), which is defined (per unit length of the cylinder) by

$$J=-\frac{e}{L}\,\mathrm{Tr}\{\hat{V}f(\hat{H}_0)\}, \tag{7}$$

where

$$\hat{V}=2B\left(-\frac{\partial}{i\partial\varphi}+\Phi\right)+\frac{\alpha}{R}\hat{\sigma}_z \tag{8}$$

is the angular velocity operator, and $f(\hat{H}_0)$ is the equilibrium density matrix ($f(\varepsilon)$ is the Fermi function). Using Eqs. (5) and (6), we can easily find the diagonal matrix elements of the operator $\hat{V}$ required to calculate the current using formula (7), whereupon we have

$$J=-\frac{eB}{2\pi R}\sum_{j,\mu}\int dk\left[2+\mu\frac{(\Lambda-1)^2}{D_j}\right]\lambda_j f(E_{j,\mu}(k)). \tag{9}$$

Quite clearly the current depends periodically on the flux with a period of one. It can also be seen that the current goes to zero for all integer and half-integer values of Φ. In fact, since the current is an odd function of Φ, J should go to zero for $\Phi=0$ (and thus for all integer values of Φ). Then, replacing j with $-(j+1)$ in the sum (9), we can see that J also goes to zero for half-integer values of Φ. The level crossing noted above also occurs for half-integer values of Φ for $k=0$: at this point the energies of the two spin-split branches of the spectrum are the same (see Fig. 1).

We shall consider the situation of a given two-dimensional electron concentration N_s. We shall calculate the current at zero temperature for which we need to express the Fermi energy E_F in formula (8) in terms of N_s and Φ, using

$$N_s=\frac{1}{4\pi^2R^2}\sum_{j,\mu}\int dk\,\theta(E_F-E_{j,\mu}(k)), \tag{10}$$

where $\theta(x)$ is the Heaviside step function. The limits of integration with respect to k in formulas (9) and (10) are determined by the form of the dispersion curves $E_{j,\mu}(k)$ which may either have a minimum for $k=0$ or a maximum at this point and two side minima. This last situation arises for $\mu=-1$ branches where $\Lambda^2/2>|(\Lambda-1)\lambda_j|$. Although the integrals in formulas (9) and (10) are expressible in terms of elementary functions, the resulting formulas are extremely cumbersome. Therefore we shall give the results (see Fig. 2) of the numerical calculations of the equilibrium current as a function of the magnetic flux for the case when for $\Phi=0$ only the $j=\pm1/2$, $\mu=-1$ states are occupied, which requires that the condition $n_s<4(\Lambda^2+|1-\Lambda|)$ ($n_s=2N_s(\pi R)^2$ be satisfied.

If the spin–orbit coupling is neglected, the linear susceptibility, defined as J/Φ, is $-2\pi eN_s/mR$, which corresponds to diamagnetism. We draw attention to the change in the sign of the linear susceptibility for $\Phi\approx0$ caused by the spin–orbit interaction (dia–para transition). We confirmed that this transition occurs if the constant Λ is greater than some positive Λ_+ or smaller than some negative Λ_- which depend on the concentration. The curves $\Lambda_+(n_s)$ and $\Lambda_-(n_s)$ are plotted in Fig. 3. The mechanism responsible for this transition becomes clear from an analysis of the partial contributions of the various terms. In the absence of spin–orbit coupling and in the presence of a weak magnetic flux, the lower level of the system is the twofold spin-degenerate $m=0$ state, where m is the azimuthal moment. In the presence of the spin–orbit interaction this level is split (for finite Φ) into $j=-1/2$, $\mu=-1$ and $j=1/2$, $\mu=-1$ terms, with the lowest term being $j=-1/2$ (for $\Phi>0$). The contributions of these sublevels correspond to diamagnetism ($j=1/2$) and paramagnetism ($j=-1/2$) and the populations differ slightly (in the pres-

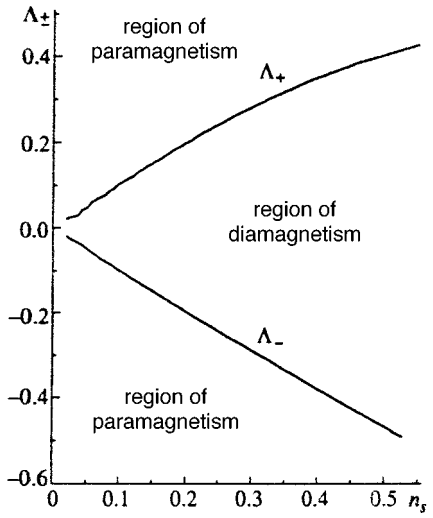

FIG. 3. Curves of $\Lambda_{\pm}(n_s)$ defining regions of diamagnetism and paramagnetism in the linear susceptibility.

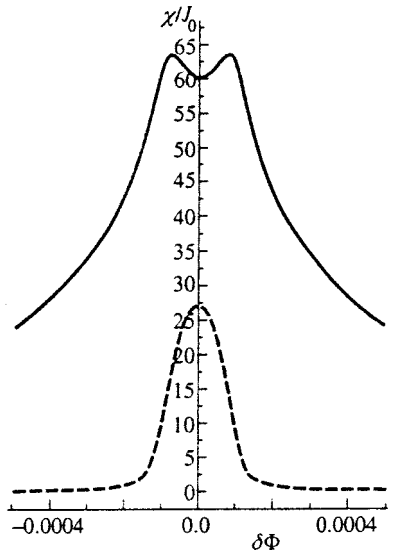

FIG. 5. Real (solid curve) and imaginary (dashed curve) parts of the dynamic susceptibility as a function of $\delta\Phi$ ($\delta\Phi=\Phi-1/2$), $\omega/B=10^{-4}$, $\omega/\nu=5$.

ence of a weak flux). In addition, electrons with the same k but different j make different contributions to the current. For $\Lambda=\Lambda_{\pm}$ the contributions to the susceptibility made by states with different $j=\pm 1/2$ cancel out.

As the flux increases, the population of the $(j=1/2, \mu=-1)$ level tends to zero, whereas the contribution of the $(j=-1/2, \mu=-1)$ level to the current decreases and at some value Φ_c changes sign. This is indicated by the behavior of the function in the integrand in Eq. (9). This behavior also explains why $J(\Phi)$ goes to zero again in Fig. 2. For $\Phi>\Phi_c$ the dependence $J(\Phi)$ is almost the same as that in the absence of the spin–orbit interaction.

In addition, near $\Phi=1/2$ the curve $J(\Phi)$ has kinks which are attributed to an abrupt change in the limits of integration with respect to k (a transition from the case of two lateral minima in $E_{-1/2,-1}(k)$ to one central one; see Fig. 4).

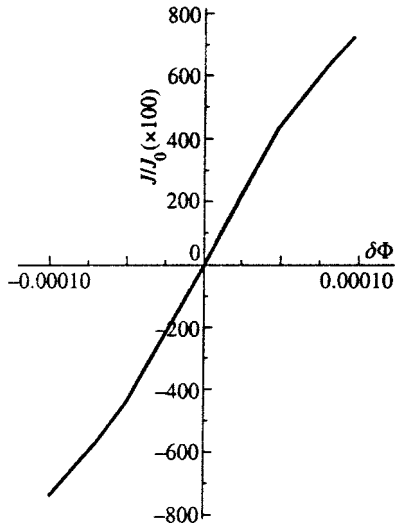

FIG. 4. Current versus Φ near $\Phi=1/2$.

3. DISPERSION OF THE MAGNETIC SUSCEPTIBILITY OF A NANOTUBE

In this section we shall examine the response of electrons situated at the surface of a cylinder to a varying magnetic flux. We shall assume that in addition to a constant flux Φ passing through the cylinder, there is also a small variable flux $\phi(t)$. We are interested in the response of the system in the linear approximation with respect to $\phi(t)$. The Hamiltonian of the system has an additional term $\hat{F}(t)$:

$$\hat{H}=\hat{H}_0+\hat{F}(t), \tag{11}$$

where $\hat{F}(t)=\phi(t)\hat{V}$. For the dynamic susceptibility $\chi(\omega)$ ($\tilde{J}_{\omega}=\chi(\omega)\phi_{\omega}$, where $\tilde{J}_{\omega}$ and ϕ_{ω} are the Fourier components of the variable components of the current and the magnetic flux) we can easily obtain the Kubo formula

$$\chi(\omega)=\frac{ie}{L}\int_{-\infty}^{\infty} dt\,\exp\{(\delta-i\omega)t\}\mathrm{Tr}\{\hat{V}\exp(i\hat{H}_0t)$$
$$\times[\hat{V},\hat{f}]\exp(-i\hat{H}_0t)\}-\frac{2e}{L}\,\mathrm{Tr}(\hat{f}). \tag{12}$$

We shall calculate the trace in formula (12) using the wave function basis (5) and (6). For off-diagonal (with respect to μ) matrix elements of the velocity operator $\hat{V}$ we have

$$V_{j,k,\mu;j,k,\bar{\mu}}=B(1-\Lambda)\frac{i\mu\Lambda k}{D_j} \tag{13}$$

($\bar{\mu}=-\mu$). As a result for $\chi(\omega)$ we find

$$\chi(\omega)=\frac{ieB^2\Lambda^2(1-\Lambda)^2}{2\pi R}$$
$$\times\sum_{j,\mu}\int dk\frac{k^2}{D_j^2}\frac{1}{\nu+i(E_{j,\bar{\mu}}(k)-E_{j,\mu}(k)-\omega)}$$
$$-\frac{e}{\pi R}\sum_{j,\mu}\int dk\, f((E_{j,\mu}(k))). \quad (14)$$

Here we have replaced the infinitely small δ from formula (12) with the phenomenological relaxation frequency ν.

Figure 5 gives results of numerical calculations of the real and imaginary parts of the susceptibility near the point $\Phi=1/2$. Quite clearly anomalies of the magnetic susceptibility should be observed in this range as a result of level crossing at low frequencies ω (however, the usual condition for the appearance of frequency dispersion $\omega\geqslant\nu$ should clearly be satisfied). The range of near-zero Φ is of no interest for this particular case of low frequencies $\omega\ll B$. It should be stressed that dispersion of the magnetic susceptibility only exists in the presence of the spin–orbit interaction, since for $\Lambda=0$, $\chi(\omega)$ is reduced to a constant, as can be seen from Eq. (12).

To conclude, allowance for spin–orbit coupling significantly alters the magnetic properties of a nanotube in a longitudinal magnetic field: in weak fields diamagnetic behavior is replaced by paramagnetic but near $\Phi=1/2$ kinks appear on the curve $J(\Phi)$ at zero temperature. Crossing of levels at $\Phi=1/2$ gives rise to anomalies in the behavior of the dynamic susceptibility at very low frequencies $\omega\ll B$.

This work was supported financially by the Russian Fund for Fundamental Research (Project No. 96-02-19058), by the program ''Physics of Solid-State Nanostructures,'' and also by the INTAS foundation (Grant No. 95-0657).

*)E-mail: chaplik@isp.nsc.ru

[1] D. Yu. Sharvin and Yu. V. Sharvin JETP Lett. **34**, 272 (1981).
[2] B. L. Al'tshuler, A. G. Aronov, and B. Z. Spivak, JETP Lett. **33**, 94 (1981).
[3] I. O. Kulik, JETP Lett. **11**, 275 (1970).
[4] V. Ya. Prinz, V. A. Seleznev, V. A. Samoylov, and A. K. Grutakovsky, Microelectron. Eng. **30**, 439 (1996); V. Ya. Prinz, V. A. Seleznev, and A. K. Grutakovsky, in *Abstracts of the 24th International Conference on Semiconductor Physics*, Jerusalem, Israel, 1998.
[5] Yu. A. Bychkov and É. I. Rashba, JETP Lett. **39**, 78 (1984)];E. I. Rashba and V. I. Sheka, in *Landau Level Spectroscopy*, edited by G. Landwehr and E. I. Rashba (Elsevier, New York, 1991), p. 178.
[6] L. I. Magarill, D. A. Romanov, and A. V. Chaplik, JETP Lett. **64**, 460 (1996)]; Zh. Éksp. Teor. Fiz. **113**, 1411 (1998) [JETP **86**, 771 (1998)]; A. V. Chaplik, D. A. Romanov, and L. I. Magarill, Superlattices Microstruct. **23**, 1231 (1998).

Translated by R. M. Durham

Thermally activated conductivity and current–voltage characteristic of dielectric phase in granular metals

E. Z. Meĭlikhov*)

Russian Research Center ''Kurchatov Institute,'' Institute of Molecular Physics, 123182 Moscow, Russia
(Submitted 11 November 1998)
Zh. Éksp. Teor. Fiz. **115**, 1484–1496 (April 1999)

Models of thermally activated linear and high-field nonlinear conductivity of a dielectric phase in granular metals (nanocomposites), i.e., aggregates of small metallic grains in a dielectric matrix, have been suggested. Given a sufficiently large spread of grain sizes, the temperature dependence of the nanocomposite conductivity should be described by a universal ''power-1/2'' law: $G \propto \exp[-(T_0/T)^{1/2}]$. An analytical expression for T_0 has been obtained. It is found that there are two regimes of nonlinear conductivity in a high electric field, namely, a low-field regime, when both the number and mobility of carriers change with the field strength, and a high-field regime, when only the mobility of carriers is variable. Analytical expressions for the nonlinear conductance in both regimes have been obtained. © *1999 American Institute of Physics.* [S1063-7761(99)02304-5]

1. INTRODUCTION

Granular metals (metal–insulator nanocomposites) are structures composed of small (with sizes $a=1-100$ nm) metallic particles embedded in dielectric matrices. They have a number of unique properties controlled by the volume fraction x of the conducting phase.[1–3] There is a certain critical point x_c such that for $x>x_c$ the material has metallic properties, whereas for $x<x_c$ it behaves like an insulator, i.e., its conductivity is thermally activated. From both physical and practical viewpoints, the most interesting properties of such materials are features of the metal–insulator transition at $x=x_c$ and various effects in nanocomposites in the dielectric phase ($x<x_c$).

This paper considers features of two such effects that have been extensively discussed in the literature, namely, the thermally activated hopping conductivity and strongly nonlinear current–voltage characteristics of dielectric nanocomposites. Although the character and magnitude of these effects are essentially controlled by material structures, they are inherent in almost all nanocomposites.

It is a well established fact that the conductivity mechanism in such systems is associated with tunneling of charge carriers between grains, and in this context it is similar to the hopping impurity conductivity in doped semiconductors.[4] Moreover, since the distances between grains vary over a wide range, it is natural to expect that their conductivity should be described by $G(T) \propto \exp[-(T_0/T)^{1/4}]$, the well-known Mott formula for the variable-range hopping conductivity.[4] Numerous experiments performed on nanocomposites of various compositions have revealed that the conductivity $G(T)$ of such systems is described by the universal ''power-1/2'' law, $G(T) \propto \exp[-(T_0/T)^{1/2}]$, where T_0 is a temperature parameter which strongly depends on x and tends to zero as $x \to x_c$. Attempts were therefore made to modify the Mott model by postulating some selection rules for allowed hops (for example, introducing a relation between the hopping length and effective grain sizes[5] in nanocomposites composed of grains with different sizes) with a view to deriving the power-1/2 law from a modified model. All these selection rules, however, seem rather far-fetched and have raised serious doubts.[2]

The power-1/2 law has been known in the theory of hopping conductivity in semiconductors for a long time. It is usually interpreted in terms of the Coulomb gap in the density of states of electrons localized at impurities, when the density of states goes to zero near the Fermi level E_F following the law $g(E) \propto (E-E_F)^2$. This issue has been discussed in numerous publications investigating, on the one hand, the nature of the Coulomb gap in the electron density of states in nanocomposite granules,[6] and on the other hand, the effect of such a gap on the temperature dependence of conductivity.[2,3] It is clear that, even if the gap scenario were realized, it would apply only to the range of low temperatures, where the gap is not smeared by thermally generated excitations (in reality, below $T<(1-10$ K). At higher temperatures the Coulomb gap cannot play an important role, and the problem of the thermally activated conductivity in granular metals remained unresolved. This paper is dedicated to this problem and shows that the power-1/2 law does not need any artificially introduced selection rules for tunneling transitions between grains, but is a direct consequence of the large spread of grain sizes typical of real nanocomposites. Moreover, the suggested simple model allows us to interpret the nonlinear conductivity of such materials in high electric fields.

2. THERMALLY ACTIVATED HOPPING CONDUCTIVITY: POWER-1/2 LAW

At zero temperature ($T=0$) and in the absence of an external electric field ($E=0$), all metallic grains in the dielectric phase of a nanocomposite are neutral, since the electrostatic energy W of any configuration of charged grains is positive.[1)] At finite temperatures, however, a thermodynamic

equilibrium of a different sort is established owing to tunneling transitions of electrons between grains, so that a fraction of grains acquire positive electric charge (and an equal fraction of grains are charged negatively). As a result, conditions are created for tunneling conductivity due to tunneling transitions (hops) from charged metallic grains to neutral particles. Since the number of charged grains increases with temperature, a rise in the conductivity should be also expected. Thus, the conductivity of dielectric nanocomposites is related to the thermal activation of carriers, and the problem is reduced to calculation of the temperature dependence of this conductivity.

If the fraction of charged grains is small, the change w_{ab} in the system energy caused by an electronic transition between two initially neutral grains is a function of the sizes a and b of these grains and separation ℓ_{ab} between them:

$$w_{ab}=e^2/2C(a,b,\ell_{ab}),$$

where

$$\frac{1}{C}=\frac{C_a+2C_{ab}+C_b}{C_aC_b-C_{ab}^2},$$

$C_a>0$ and $C_b>0$ are the capacitance coefficients of grains a and b, and $C_{ab}<0$ is their electrostatic induction factor.[7] In the limit $\ell_{ab}\gg a,b$

$$C_a\approx\varepsilon\frac{a}{2}\left(1+\frac{ab}{\ell_{ab}^2}\right)\approx\varepsilon a/2,$$

$$C_b\approx\varepsilon\frac{b}{2}\left(1+\frac{ab}{\ell_{ab}^2}\right)\approx\varepsilon b/2,\quad -C_{ab}\approx\varepsilon\frac{ab}{\ell_{ab}}\ll C_a,C_b,$$

where ε is the dielectric constant of the material. Thus, the electrostatic energy satisfies

$$w_{ab}\approx\frac{e^2}{2\varepsilon}\left(\frac{1}{a}+\frac{1}{b}\right)$$

and in the limit $a\ll b$ it is described by a simple formula

$$w_{ab}\approx(e^2/\varepsilon a).$$

However, the latter relation applies only under the condition that the host material containing the metallic grains is insulating and does not screen their electrostatic fields. In reality, the system in question is conducting owing to the tunneling transitions of electrons or holes from charged to neutral grains and can be characterized by a finite screening length L. Hence

$$w_{ab}\approx\frac{e^2}{\varepsilon}\left(\frac{1}{a}-\frac{1}{L}\right).$$

The mean thickness of the tunneling barrier for such transitions depends on the mean grain size and goes to zero when the latter reaches the so-called percolation radius L_c, at which the grains would form an infinite cluster[4] and which is related to the volume fraction x of metallic grains by the simple formula $(L_c/a)^3=x_c/x$. If the percolation cluster contains mostly grains of one size a (see below), it is natural to set $L=L_c$ (Ref. 8) and express the electrostatic energy by the formula

$$w_a\approx\frac{e^2}{\varepsilon a}\left[1-\left(\frac{x}{x_c}\right)^{1/3}\right]. \tag{1}$$

If all grains were of the same size a, the concentration of charged grains with both charge signs (their charges being $\pm e$) would be $N_\pm=N\exp(-w_a/kT)$, where $N=x/(4\pi a^3/3)$ is the number of grains per unit volume. (For $N_\pm\ll N$ such a system is similar to a weakly compensated system of donors and acceptors in a semiconductor at zero temperature.[4] The role of charged donors and acceptors in the system in question is played by negatively and positively charged grains.) In this case, the resistance of the material is

$$R\propto\frac{1}{N_\pm}\exp\left(\frac{2\langle\ell\rangle}{\lambda}\right)\propto\exp\left(\frac{w_a}{kT}+\frac{2\langle\ell\rangle}{\lambda}\right)\propto\exp\left(\frac{w_a}{kT}\right). \tag{2}$$

Here $\langle\ell\rangle$ is the mean separation between grains, $\lambda\sim\hbar/(mW)^{1/2}$ is the electron wavelength in the insulator, and W is the tunneling barrier height (which is close to half the dielectric band gap).

The resulting temperature dependence $\rho(T)$ is markedly different from the power-1/2 law because the assumption of an equal size for all grains is unrealistic. In fact, with the technique used in fabrication of nanocomposites the sizes of metallic grains vary over a fairly wide range. Charge transfer between grains of different sizes (from a charged grain of size a to a neutral grain of size b) requires a certain amount of energy:

$$w_{ab}\approx\frac{e^2}{\varepsilon}\left(\frac{1}{b}-\frac{1}{a}\right)\left[1-\left(\frac{x}{x_c}\right)^{1/3}\right],$$

which can be offset by absorbing a phonon (on the other hand, a transition from a small grain to a larger one may generate a phonon). Given the wide spread of grain sizes, it is natural to expect a large spread of activation energies w_{ab} of different hops.

This model is very similar to the percolation model of variable-range hopping conductivity in the theory of semiconductors, one of whose tenets is the large spread of energies of the states participating in electronic transitions.[4] According to this model, there is an optimal (temperature-dependent) hopping length ℓ_{ab} between grains a and b, which is determined by the interplay between the tunneling probability proportional to $\exp(-2\ell_{ab}/\lambda)$ and the probability of thermal activation over the barrier w_{ab}, which is proportional to $\exp(-w_{ab}/kT)$.

If the distribution function $f(a)$ of grain sizes is known (for example, the exponential distribution function $f(a)=(1/a_0)\exp(-a/a_0)$ is quite common[9]), the concentration of charged grains with sizes close to a is

$$N_a\propto f(a)\exp\left[-\frac{w(a)}{kT}\right]. \tag{3}$$

Then the fraction of the system resistance due to charged grains with size a is[2)] $R_{ab}\propto(1/N_a)\rho_{ab}$, where $\rho_{ab}\propto\exp[2\ell_{ab}/\lambda+w_{ab}/kT]$. Thus,

$$R_{ab}\propto\exp\left[\frac{2\ell_{ab}}{\lambda}-\ln f(a)+\frac{w_a+w_{ab}}{kT}\right]$$
$$\propto\exp\left[\frac{2\ell_{ab}}{\lambda}-\ln f(a)+\frac{w_b}{kT}\right]. \quad (4)$$

The hopping length ℓ_{ab} equals the separation between grains a and b. For $f'(a)<0$ the optimal hops are those between grains with close sizes (in this specific case a). In fact, hops to more widespread smaller grains ($b<a$) would require a higher activation energy, $w(b)>w(a)$, and hops to less common grains of larger sizes, $b>a$, would be considerably longer. Both would give rise to a higher resistivity. Hence we set $b=a$ in Eq. (4). Then $\ell_{ab}=\alpha(T)[f(a)]^{-1/3}$, where α is independent of a. Thus,

$$R_{ab}\propto\exp\left[\alpha(T)[f(a)]^{-1/3}-\ln f(a)+\frac{\beta}{aT}\right], \quad (5)$$

where $\beta=(e^2/\varepsilon k)[1-(x/x_c)^{1/3}]$.

The resistance of a separate current path connecting two opposite faces of a sample is determined by the product of the resistivity ρ_{ab} corresponding to one hop times the number of such hops along this path, which is inversely proportional to the hopping length ℓ_{ab}. The total sample resistance R is inversely proportional to the number of current paths, which is, in turn, inversely proportional to ℓ_{ab}. Thus, the total resistance satisfies $R\propto R_{ab}$ and is controlled by hops between optimal grains, whose size is calculated by minimizing the argument of the exponential function in Eq. (5) through variation of a. If the spread of actual grain sizes is small, the functions $[f(a)]^{-1/3}$ and $\ln f(a)$ can be approximated by linear expressions around the optimal size $a_{\rm opt}$ [unless the distribution function $f(a)$ has a peculiar shape]:

$$[f(a)]^{-1/3}\approx\text{const}+\gamma a, \quad \ln f(a)\approx\text{const}-\gamma' a,$$

where

$$\gamma=-\frac{1}{3}[f(a_{\rm opt})]^{-4/3}f'(a_{\rm opt}), \quad \gamma'=-\frac{f'(a_{\rm opt})}{f(a_{\rm opt})}.$$

Then

$$R(a)\propto\exp\left(\tilde{\alpha}a+\frac{\beta}{aT}\right), \quad (6)$$

where $\tilde{\alpha}=\gamma\alpha+\gamma'$.

Let $f'(a_{\rm opt})<0$ and the temperature dependence of $\tilde{\alpha}$ be so weak that it could be neglected (this assumption will be justified below). Then $\tilde{\alpha}>0$ and minimization of the exponent in Eq. (5) by varying a yields the optimal grain size $a_{\rm opt}\propto T^{-1/2}$, i.e.,

$$R\propto\exp\left[\left(\frac{T_0}{T}\right)^{1/2}\right], \quad (7)$$

where T_0 is the temperature parameter that controls the system resistance. This is the power-1/2 law, which is often mentioned in publications dedicated to nanocomposite properties.

Note that Eq. (7) was derived using linear approximations of functions $[f(a)]^{-1/3}$ and $\ln f(a)$. In cases when these approximations cannot be applied, the temperature dependence of the resistance should be different. For example, in systems with uniform distributions of grain sizes [$f(a)=\text{const}$ for $a_{\min}<a<a_{\max}$], $f'(a)=0$, and optimal hops are those between the smallest grains. In this case we have $\rho\propto\exp(w/kT)$, where $w=e^2/\varepsilon a_{\max}[1-(x/x_c)^{1/3}]$.

Thus, we have shown that the power-1/2 law applies to thermally activated conductivity of nanocomposites under the following conditions:

(1) $\tilde{\alpha}(T)$ is a weak function of temperature;

(2) the relative spread of relevant grain sizes is not too large;

(3) the distribution function of grain sizes is such that $f'(a_{\rm opt})<0$, i.e., the number of grains is a dropping function of their size.

Whereas the latter condition is usually satisfied under real fabrication conditions, the applicability of the former two must be tested. With this end in view and in order to estimate the parameter T_0 of the suggested model, let us clarify the meaning of $\tilde{\alpha}$ in Eq. (6).

As follows from the derivation of Eq. (5), this parameter is related to the concentration of grains with the optimal size $a_{\rm opt}$. It is clear that such grains are those whose sizes belong to a certain interval $a_{\rm opt}-\Delta a<a<a_{\rm opt}+\Delta a$, and the width of this interval is determined by the condition that grains with sizes outside this interval make a small contribution to the conductivity. Let us transform Eq. (6) to $\rho\propto\exp[\xi(a)]$, where $\xi(a)=\tilde{\alpha}a+1/a\tau$, $\tau=T/\beta$, and (in a rough approximation) $\tilde{\alpha}$ is independent of temperature. The minimal resistivity (which controls the total sample conductivity) is due to the grains with size $a=a_{\rm opt}=(1/\tilde{\alpha}\tau)^{1/2}$ and is $\rho(a_{\rm opt})\propto\exp\xi_{\rm opt}$, where $\xi_{\rm opt}=\xi(a_{\rm opt})=2(\tilde{\alpha}/\tau)^{1/2}$. Note that, as in all percolation models (specifically, in the hopping conductivity model[4]), $\xi_{\rm opt}\gg 1$. Since the resistivity is an exponential function of ξ, we assume that the optimal grains are those whose ξ is within unity from the optimal value $\xi_{\rm opt}$. The function $\xi(a)$ can be approximated near the optimal size $a=a_{\rm opt}$ as

$$\xi(a)\approx\xi_{\rm opt}+(1/\tau a_{\rm opt}^3)(a-a_{\rm opt})^2.$$

By setting $\xi(a)-\xi(a_{\rm opt})=1$, we obtain $(\Delta a)^2-a_{\rm opt}^3\tau=2a_{\rm opt}^2/\xi_{\rm opt}$. This implies $(\Delta a/a_{\rm opt})^2=2/\xi_{\rm opt}\ll 1$, i.e., the relative spread of optimal grain sizes is really small.

Further, for definiteness let us analyze a system with an exponential distribution function of grain sizes: $f(a)=(1/a_0)\exp(-a/a_0)$, Ref. 9. Let the total number of grains of all sizes per unit volume be N. Obviously,[3)]

$$N=\frac{x}{\int(4\pi/3)a^3f(a)da}=\frac{x}{8\pi a_0^3}.$$

Then the density of efficient grains with sizes within the range defined above is

$$N_{\rm opt}=Nf(a)\cdot 2\Delta a=2N(a_{\rm opt}^3\tau)^{1/2}\frac{1}{a_0}\exp(-a_{\rm opt}/a_0).$$

The average separation between them is

$$\gamma_{ab} \approx \frac{1}{4N_{\rm opt}^{1/3}} \approx \gamma_0 \exp\left(\frac{a_{\rm opt}}{3a_0}\right),$$

where

$$\gamma_0 = \frac{1}{3}\left(\frac{a_0}{N}\right)^{1/3} (\tau a_{\rm opt}^3)^{-1/6} \propto T^{1/12}.$$

In what follows we will neglect the weak temperature dependence of γ_0 and for $a_{\rm opt} \leqslant a_0$ take $\gamma_{ab} \approx {\rm const} + \tilde{\gamma}_{ab}$, where (in the linear approximation) $\tilde{\gamma}_{ab} = (\gamma_0/3a_0)a_{\rm opt}$. In this approximation $\ln f(a) \approx {\rm const} - a_{\rm opt}/a_0$. In the derivivation of the optimal grain size from Eq. (6), only the temperature dependent terms $2\tilde{\gamma}_{ab}/\lambda$ and $a_{\rm opt}/a_0$ are important, whose sum $(2\gamma_0/3\lambda + 1)(a_{\rm opt}/a_0)$ is included in the definition of $\tilde{\alpha}$ given above.

Under the conditions of tunneling conductivity,[4] $\gamma_0/\lambda \gg 1$ holds, so $\tilde{\alpha} \approx 2\gamma_0/3\lambda a_0$. Using the expression for $a_{\rm opt}$ derived in the rough approximation [i.e., neglecting the temperature dependence of $\tilde{\alpha}(T)$], we obtain in the next approximation

$$\tilde{\alpha} \approx (\lambda N^{1/3})^{-4/3} \tau^{1/9}/2a_0^{8/9} \propto T^{1/9},$$

i.e., although this parameter is a function of temperature, this dependence is very weak and cannot notably change the power-1/2 law. At the same time, the explicit expression for $\tilde{\alpha}$ yields both the optimal grain size $a_{\rm opt}$ and the temperature parameter T_0 in Eq. (7). To this end, it suffices to insert the more accurate expression for $\tilde{\alpha}$ in the approximate expressions $a_{\rm opt} = (1/\tilde{\alpha}\tau)^{1/2}$ and $\xi_{\rm opt} = (2\tilde{\alpha}/\tau)^{1/2}$. As a result, the power-1/2 law is slightly modified and takes the form

$$\rho \propto \exp\left[\left(\frac{T_0}{T}\right)^{4/9}\right], \tag{8}$$

where

$$kT_0 \approx \left(\frac{e^2}{\varepsilon a_0}\right)\left(\frac{a_0}{\lambda}\right)^{3/2} \psi(x), \quad \psi(x) = x^{-1/2}\left[1 - \left(\frac{x}{x_c}\right)^{1/3}\right]. \tag{9}$$

Thus, the optimal grain size is

$$a_{\rm opt} = a_0\left(\frac{x}{4\pi}\right)^{1/2}\left(\frac{\lambda}{a_0}\right)\left(\frac{T_0}{T}\right)^{5/9}, \tag{10}$$

i.e., it grows with decreasing temperature and drops as the temperature rises. This defines natural bounds for the applicability of the model: it is valid as long as $a_{\rm min} < a_{\rm opt} < a_{\rm max}$, i.e., in the temperature range

$$T_{\rm min} < T < T_{\rm max},$$

$$T_{\binom{\rm min}{\rm max}} = (4\pi)^{9/10} \frac{e^2}{k\varepsilon a_0}\left[\frac{a_0}{a_{\binom{\rm max}{\rm min}}}\right]^{3/2} \times \left[\frac{\lambda}{a_{\binom{\rm max}{\rm min}}}\right]^{3/10} x^{-7/5}\left[1 - \left(\frac{x}{x_c}\right)^{1/3}\right]. \tag{11}$$

For typical values $\lambda = 2 \cdot 10^{-8}$ cm, $a_0 = 10^{-7}$ cm, $\varepsilon = 5$, and $x_c = 0.5$, we find that, for example, at $x/x_c = 0.3$, the model (hence the power-1/2 law) is valid in a wide temperature range 6 K $< T <$ 500 K. But this range narrows and shifts to lower temperatures as $x \to x_c$. This means that at higher temperatures $a_{\rm opt} = a_{\rm max}$ holds everywhere. It is natural to expect a transition to the conductivity behavior typical of nanocomposites with grains of equal sizes and $G \propto \exp(-T_0/T)$. Such a change in behavior was observed in experiments with Al–Ge granular films[10] as x varied over the interval[5] $0.3 < x < 0.45$.

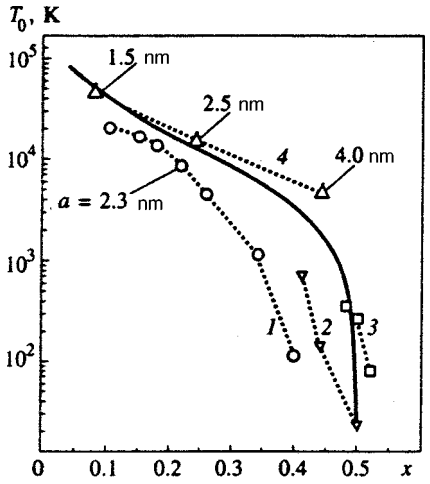

FIG. 1. Temperature parameter $T_0(x)$ as a function of metal content for nanocomposites fabricated from different materials. Experimental data were taken from the literature as follows: (*1*) W–Al_2O_3 (Ref. 2); (*2*) Fe–SiO_2 (Ref. 11); (*3*) Ni–SiO_2 (Ref. 12); (*4*) Ni–SiO_2 (Ref. 5). The solid curve plots calculations by Eq. (9) with parameters given in the text. Average sizes of metallic grains are given.

By applying the same technique to systems with other distribution functions, one can obtain similar results.

In comparing the results deriving from Eq. (9) to experimental data, one should keep in mind two complications. The first is the poor accuracy of the metal fraction x in a nanocomposite measured in experiments, which is important for x close to x_c. The second is the correlation between the average grain size a_0 and the volume fraction of a metal determined by technological conditions. In reality, a_0 increases with x (so, in Ni–SiO_2 systems[5] we have $a_0 = 15$, 25, and 40 Å at $x = 0.08$, 0.24, and 0.44, respectively). For this reason, one can expect good agreement between experimental data and calculations only in the overall shape of the temperature dependence of conductivity (the power-1/2 law). As concerns the absolute value of T_0 and its dependence on the metal content, $T_0(x)$, the theory could be deemed successful if it predicted T_0 correctly within one order of magnitude and the overall shape of the $T_0(x)$ curve. As for the temperature dependence of the resistivity, it was shown in dozens of experiments performed with different nanocomposite systems in a wide temperature range (from room to liquid-helium temperatures) that it follows the law $\rho(T) \propto \exp[(T_0/T)^{1/2}]$ (see, for example, the reviews[1,2]). Figure 1 plots experimental data by different authors for the function $T_0(x)$ in nanocomposites manufactured from different materials (the large spread of experimental data is caused by the complications mentioned above). This graph also shows a theoretical curve of $T_0(x)$ calculated by Eq. (9) with the

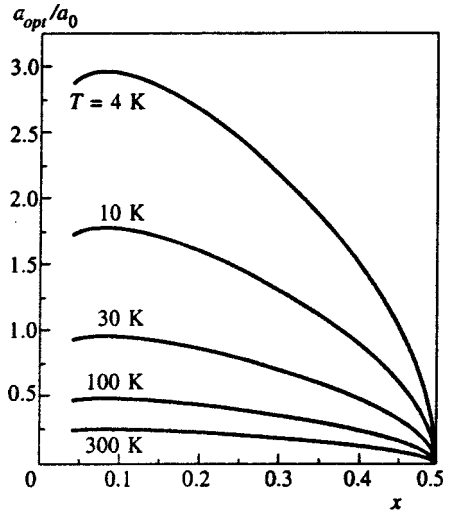

FIG. 2. Optimal size of metallic grains in a nanocomposite that make the major contribution to the conductivity. The calculations were performed using Eq. (10) with parameters given in the text.

following parameters: $\lambda = 2\times 10^{-8}$ cm, $a_0 = 10^{-7}$ cm, $\varepsilon = 5$, and $x_c = 0.5$. The agreement between the experimental data and calculations is satisfactory.

It was noted above that the temperature dependence of the optimal grain size is approximately described by the simple formula $a_{\rm opt} \propto T^{-1/2}$. Its dependence on the metal fraction, $a_{\rm opt}(x)$, is also easily understandable. The mean separation between grains decreases with increasing x, and small grains (characterized by high activation energies) also contribute to the conductivity. This should yield a smaller $a_{\rm opt}$. Calculations plotted in Fig. 2 indicate that the optimal grain size is smaller than their average size in a wide region of x and T. For nanocomposites with $x/x_c \geqslant 0.9$ this condition ($a_{\rm opt} \leqslant a_0$) is satisfied over the range down to the liquid-helium temperature.

To conclude this section, note that the power-1/2 law was derived by Sheng *et al.*[5] under the assumption that there is a structural relation between the grain size a and separation ℓ between them: $\ell/a = \text{const}$. Adkins,[2] however, correctly noted that the assumption had not been supported by structural analysis data. It follows from our analysis that the power-1/2 law should be valid under a less restrictive condition $\partial\ell/\partial a = \text{const}$ near $a = a_{\rm opt}$ (the condition $\ell/a = \text{const}$ is a very special case of our condition). Note also that some of the sophisticated methods used by different authors to justify the power-1/2 law can be boiled down (although it is not obvious sometimes) to linearizing the actual portion of function $[f(a)]^{-1/3}$. For example, Sheng[3] linearized the distribution function $\phi(w_{ij})$ of the activation energy in the region of low energies (in fact, he set $\phi(w_{ij}) \propto w_{ij}$ for $w_{ij} < w_0$), then a numerical calculation led to the power-1/2 law. This result, however, was valid only in the temperature range $kT \leqslant w_0$. It is not surprising because (given $w_{ij} \propto 1/a$) the postulated function $\phi(w_{ij})$ corresponded to the distribution function of grain sizes $f(a) = \phi[w_{ij}(a)](dw_{ij}/da)^{-1} \propto 1/a^3$, which implies $f^{-1/3}(a) \propto a$, the condition mentioned above (moreover, it is the function $f(a) \propto 1/a^3$ that satisfies the condition $\ell/a = \text{const}$ discussed above). Nonetheless, the coincidence between these numerical calculations of the temperature dependence of the conductivity and the result of our analysis provides evidence in favor of the suggested approach.

3. CURRENT–VOLTAGE CHARACTERISTIC

According to the model described in the previous section, the conductivity in the weak-field limit is connected with tunneling transitions of charges (electrons or holes) between charged and neutral grains of a quite definite (optimal) size $a = a_{\rm opt}$, which is considerably larger than the size of the smallest close-set grains. Although the probability of tunneling between the latter is higher than between more wuidely spaced grains of the optimal size, the number of charges on small grains is limited by their high ionization energy, so their contribution to the conductivity is small. In sufficiently high electric fields, transitions between large charged grains and smaller neutral grains, whose probability is low in weak electric fields, become possible, so the conductivity is effected via more probable transitions involving closely set small grains. Both the number of carriers (owing to the change in the optimal size a_E of donor grains) and their mobility (owing to the change in the optimal size b_E of conducting grains) vary with the electric field.

In accordance with this reasoning, let us assume that an electric field of strength E offsets the activation energy w_{ab} of the transition from a grain of size a to a grain of size $b < a$, i.e.,

$$eE\ell_{ab} = w_{ab} = \frac{e^2}{\varepsilon}\left[1 - \left(\frac{x}{x_c}\right)^{1/3}\right]\left(\frac{1}{b} - \frac{1}{a}\right). \tag{12}$$

We ignore the fact that such transitions can be driven by thermal activation, i.e., absorption of phonons. This can be done if the temperature is moderate, $kT \ll w_{ab}$ (qualitative estimates of the effect of finite temperature will be given below).

Then we can set $w_{ab} = 0$ in Eq. (4) and assume that the conductivity in these conditions is due to tunneling (through a distance $\sim \ell_{bb}$) between grains of size b. Therefore the first term of the exponent in Eq. (4) equals $2\ell_{bb}/\lambda$, whereas the term related to the number of carriers is still determined by ionization energy w_a of grains with size a, but now should be expressed with due account of Eq. (12) relating a and b:

$$\frac{w_a}{kT} = \left(\frac{1}{b\tau} - \frac{\ell_{ab}E}{Q\tau}\right), \quad Q = \frac{e}{\varepsilon}\left[1 - \left(\frac{x}{x_c}\right)^{1/3}\right].$$

Since we have $a > b$, we can equate the average separation ℓ_{ab} between grains a and b to the smaller distance ℓ_{bb}. Finally, as previously, we can neglect the term proportional to $\ln f(a)$. As a result, the exponent ξ takes the form

$$\xi = \alpha_E b + \frac{1}{\tau b}, \quad \alpha_E = \tilde{\alpha}\left(1 - \frac{\lambda E}{2Q\tau}\right), \tag{13}$$

i.e., it is similar to the case of linear conductivity discussed in the previous section [Eq. (6)].

As is usual, the resistance $R\propto\exp(\xi)$ is determined by the minimal value of ξ, which is equal in this case to $\xi_E=2\xi_{\rm opt}(1-\lambda E/2Q\tau)^{1/2}$. Thus the total conductance of the sample $G=1/R$ is given by[6)]

$$G(E)\propto\exp[-2\xi_{\rm opt}(\sqrt{1-\lambda E/2Q\tau}-1)]. \tag{14}$$

The optimal grain sizes a and b increase with the electric field strength:

$$b_E=(\alpha_E\tau)^{-1/2}=a_{\rm opt}\left(1-\frac{\lambda E}{2Q\tau}\right)^{-1/2},$$

$$a_E=\frac{b_E}{1-E\not{e}_{bb}b_E/Q}=a_{\rm opt}\left[\left(1-\frac{\lambda E}{2Q\tau}\right)^{1/2}-\frac{a_{\rm opt}\not{e}_{bb}E}{Q}\right]^{-1}, \tag{15}$$

but in all fields $b_E<a_E$.

This analysis is, naturally, valid until $a_E<a_{\max}$. In higher electric fields, the size of donor grains is fixed at $a_E=a_{\max}$, and the size of conducting grains decreases: $b_E=a_{\max}/(1+a_{\max}\not{e}_{bb}E/Q)$. Further increase in the electric field strength does not affect the number of carriers supplied by grains of the maximal size, but their mobility increases owing to hops to smaller grains separated by smaller distances. In this case we have $\xi_E=\tilde{\alpha}b_E$, and the conductance is

$$G(E)\propto\exp\left[-\frac{E_0}{E+(\tilde{\alpha}a_{\max})^{-1}E_0}\right], \tag{16}$$

where $E_0=\tilde{\alpha}Q/\not{e}_{bb}\propto[1-(x/x_c)^{1/3}]$. Since $(\tilde{\alpha}a_{\max})^{-1}\sim[(a_{\max}/a_{\rm opt})\xi_{\rm opt}]^{-1}\ll 1$, it follows from Eq. (16) that even in fields $E\ll E_0$

$$G(E)\propto\exp\left(-\frac{E_0}{E}\right), \tag{17}$$

which is formally equivalent to the result obtained by Sheng *et al.*[5] for a specific distribution function $f(a)$ and under different assumptions.

The characteristic field strength E_0 can be estimated by the relation

$$E_0\sim\frac{e}{\varepsilon a_0^2}\left[1-\left(\frac{x}{x_c}\right)^{1/3}\right],$$

which yields $E_0\sim 10^6$ V/cm for $a_0\sim 3\times 10^{-7}$ cm, $x\sim 0.2$, and $x_c\sim 0.5$. This estimate is in agreement with the experimental data[5] measured for $T\sim 1$ K: $E_0=(3.5,\ 1.0,\ 0.23)\cdot 10^6$ V/cm for systems with $x=0.08$, 0.24, and 0.44, respectively. After setting $x_c=0.5$, we find that, in accordance with the expression for E_0 given above, the proportion among these fields should be 2.3:1:0.19, which is fairly close to the experimental data.

As for the low-field regime of nonlinear conductivity, it should be noted that this model predicts qualitative differences between curves of conductance versus electric field measured in the two regimes mentioned above: the curve of $G(E)$ should have a positive curvature ($G''(E)>0$) in the low-field regime and negative ($G''(E)<0$) in the high-field regime. This can be used as a criterion for distinguishing between the two regimes. The fact that $G(E)$ curves of the first type were recorded neither in fields $E\sim 10^5$–10^6 V/cm (Ref. 5) nor for $E\sim 10^3$–10^4 V/cm (Ref. 13) indicates that the boundary between these two regimes is in the range of fields $E<10^3$ V/cm. Preliminary results[7)] of detailed studies in these and lower fields are in agreement with the present model.[14]

In conclusion, let us briefly discuss the effect of finite temperature on our results. In this case, transitions $a\to b$ are driven not only by electric field, but also by phonons of energy $\sim kT$, which is equivalent, according to Eq. (12), to the replacement $E\to E-E_T$, where $E_T=kT/e\not{e}_{aa}$. Then we have instead of Eq. (14)

$$\frac{G(E)}{G(0)}\propto\begin{cases}1, & E\leqslant E_T,\\ \exp[-2\xi_{\rm opt}(\sqrt{1-\lambda(E-E_T)/2Q\tau}-1)], & E>E_T\end{cases}, \tag{18}$$

which means that there is a region of linear conductivity in low electric fields with the upper boundary $E_T\propto T$.

At higher temperatures we have $a_{\rm opt}\propto T^{-1/2}\to a_{\min}$, and an electric field is no longer necessary for driving currents through small grains. This does not mean, however, that it has no effect on the conductivity. In this case, we should reconsider the nonlinearity mechanism suggested earlier,[5] namely, the effect of electric field on the tunneling probability between grains.

Thus, on the basis of this simple model it is possible to offer a qualitative interpretation of various features of thermally activated conductivity of granular metals, including its nonlinearity in high electric fields.

The author is indebted to B. A. Aronzon and V. V. Ryl'kov for helpful discussions and information about experimental results prior to publication. The work was supported by the Russian Fund for Fundamental Research (Grants 96-02-18429-a and 98-02-17412-a) and by the joint program of Center National de Recherche Scientifique of France and Russian Fund for Fundamental Research (Grant 98-02-22037).

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

1)Here we neglect the fact that a fraction of grains in the ground state ($T=0$) are charged in the presence of the Coulomb gap.[6]

2)Hereafter we assume that only hops between nearest neighbors are important. Although the hopping length varies with the temperature (in accordance with the change in the optimal grain size), these are always hops between nearest neighbors. This is a fundamental departure from the Mott picture of the variable-range hopping conductivity and is related to the purely geometrical consideration that for $x>x_c/10$ grains of optimal sizes are shielded by their nearest (or second nearest) neighbors from their farther neighbors. This makes impossible tunneling transitions between farther neighbors, which would otherwise require a smaller energy.

3)In the calculation of N the integration should be performed over the interval $a_{\min}<a<a_{\max}$, where $a_{\min}$ and $a_{\max}$ are the maximal and minimal grain size, respectively. If $a_{\max}\geqslant 3a_0$ and $a_{\min}\leqslant a_0$, the result differs from that given in the text by only several percent.

4)This means, in particular, that the term proportional to $\ln f(a)$ in Eqs. (4) and (5) can be neglected.

[5)]At higher x a transition to a nonactivated conductivity $G(T)$ took place. The most probable cause of this change is a more complicated structure of the metallic component for x approaching x_c. In this case the simple model of isolated ball-shaped particles of different sizes no longer applies.
[6)]No complications occur when the radicand in Eq. (14) goes to zero (see below).
[7)]The experiments were performed on samples with $x \approx x_c$ in fields $E = 10^2 - 10^3$ V/cm. The conductance remained constant in the region of low electric fields, whose width increased with temperature, and then changed by an order of magnitude.

[1] See, e.g., the special issue of Phil. Mag. B **65** (1992).
[2] C. J. Adkins, in *Metal–Insulator Transitions Revisited*, ed. by P. P. Edwards and C. N. R. Rao, Taylor & Francis (1995); J. Phys.: Condens. Matter **1**, 1253 (1989).
[3] P. Sheng, Philos. Mag. B **65**, 357 (1992).
[4] B. I. Shklovskii and A. L. Efros, *Electronic Properties of Doped Semiconductors*, Springer, New York (1984).
[5] P. Sheng, B. Abeles, and Y. Arie, Phys. Rev. Lett. **31**, 44 (1973).
[6] S. T. Chui, Phys. Rev. B **43**, 14274 (1991); E. Cuevas, M. Ortuño, and J. Ruiz, Phys. Rev. Lett. **71**, 1871 (1993).
[7] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon, Oxford (1980).
[8] B. A. Aronzon, A. E. Varfolomeev, D. Yu. Kovalev *et al.*, Fiz. Tverd. Tela **40**, No. 6 (1999) (in press); A. A. Likal'ter, private communication.
[9] A. B. Pakhomov, X. Yan, N. Wang *et al.*, Physica A **241**, 344 (1997).
[10] G. Eytan, R. Rosenbaum, D. S. McLachlan, and A. Albers, Phys. Rev. B **48**, 6342 (1993).
[11] B. Zhao and X. Yan, Physica A **241**, 367 (1997).
[12] J. I. Gittleman, Y. Goldstein, and S. Bozowski, Phys. Rev. B **5**, 3609 (1972).
[13] S. A. Gurevich, V. V. Khorenko, L. Yu. Kupriyanov *et al.*, JETP Lett. **64**, 736 (1996).
[14] V. V. Ryl'kov, private communication.

Translation provided by the Russian Editorial office.

Application of renormalization group techniques to transport in the presence of nonlinear sources and sinks

É. V. Teodorovich*)

Institute for Problems in Mechanics, Russian Academy of Sciences, 117526 Moscow, Russia
(Submitted 7 September 1998)
Zh. Éksp. Teor. Fiz. **115**, 1497–1510 (April 1999)

The paper considers the problem of spreading of a localized distribution due to diffusion in the presence of nonlinear sources and sinks modeling annihilation and creation of material in chemical reactions. The evolution of the parameters characterizing the problem, namely, the amplitude and radius of the distribution, has been investigated under the assumption that the distribution is self-similar. These parameters have been calculated using the renormalization-group method with renormalization of the diffusion coefficient and the total amount of material. As a result, it is possible to classify various regimes of the asymptotic behavior of the system at long times according to the sign of nonlinear interaction constant and the spatial dimensionality. In particular, the conditions under which the regime of asymptotic freedom and the peaking regime are realized have been found. The renormalization group method not only allows us to calculate the exponents of functions with power-law behavior, but also to track the transition to the asymptotic regime and calculate numerical coefficients. © *1999 American Institute of Physics.* [S1063-7761(99)02404-X]

1. INTRODUCTION

The problem of transport in the presence of nonlinear sources and sinks has long been long been at the forefront of research because this phenomenon has implications for physics, chemistry, engineering, biology, ecology, and other sciences. The theoretical description of these phenomena is based on a quasilinear parabolic equation of the form

$$\left[\frac{\partial}{\partial t}-D_0\Delta\right]C(\mathbf{r},t)+\lambda C^{1+2\delta}(\mathbf{r},t)=0. \tag{1.1}$$

In particular, Eq. (1.1) occurs in modeling processes of chemical kinetics, coagulation of aerosol particles, filtering in porous media, heat transfer with due account of losses and heating due to radiation, etc. To be more specific, in what follows we will relate our analysis to chemical kinetics,[1] so the quantity $C(\mathbf{r},t)$ will be called the concentration of a substance, D_0 the molecular diffusion coefficient, $n=1+2\delta$ the order of the chemical reaction, and λ the reaction rate constant. The case $\lambda>0$ corresponds to annihilation of a substance in the course of the reaction, and $\lambda<0$ to creation. Studyies of the long-time asymptotics of the solution reveal that there is a critical spatial dimension $d_c=1/\delta$ for absorption reactions above which diffusion processes should be unimportant, the solution as a function of time depends on the order of the chemical reaction as $C(\mathbf{r},t)\propto(\lambda t)^{-1/2\delta}$ and appears to be independent of the initial substance concentration in agreement with the results of the mean-field approximation.

However, for dimensions below the critical value a specific nonlinear regime is observed and the concentration decays asymptotically with time as $C(\mathbf{r},t)\propto t^{-d/2}$, which implies that spatial fluctuations dominate in the concentration. A convenient technique for investigating fluctuations in a system with an infinite number of essential degrees of freedom is an application of quantum field theory methods. According to the Martin–Siggia–Rose formalism of ''field doubling,''[2] a classical system described by Eq. (1.1) is equivalent to a certain quantum system in the sense that they have identical characteristic functionals (functional Fourier transforms of the probability density identified with the concentration) expressed in the form of path integrals for two fields presented by creation and annihilation operators of particles involved in a reaction.[3]

The technique of renormalized perturbation theory borrowed from the quantum field theory was employed in investigating equation system (1.1) by Doi.[4] Peliti[3] showed that for binary reactions ($n=2$, $\delta=1/2$) one can sum up the perturbation series, calculate exact renormalizations of the propagator and vertex, and determine the exponent in the power function of time, which appears to be valid for all orders of perturbation theory. The idea of using the renormalization group (RG) method in summing the infinite perturbation series was suggested and exploited by Ohtsuki.[5] He investigated the problem of the evolution of a given initial distribution for $d<d_c$ and, under the assumption that the asymptotic form is independent of the initial distribution, obtained damping with exponent $-d/2$. Like Peliti,[3] Ohtsuki[5] analyzed only the renormalization of reaction rates (vertices), and no renormalization of field amplitudes was needed. The application of the RG technique allowed them to determine not only the exponents but also the numerical coefficients of the power functions, and describe the transition to the asymptotic regime as well. Lee[6] generalized this analysis to the case of a reaction of arbitrary order (unlike the case of binary reactions analyzed previously), and in addition, he calculated numerical amplitude coefficients.

This paper suggests a slightly different method based on the RG technique for investigating the asymptotic distribution function, which allows one to perform an analytic continuation with respect to the order of the chemical reaction as well as the spatial dimension. This is rather difficult in the field approach, since the reaction order governs the topological structure of the Feynman diagrams. Moreover, if the quantum field theory formalism is not used when the Lagrangian structures for creation and annihilation reactions are different, these processes are described similarly in our approach, and the only difference is in the sign of the reaction constant λ.

2. PROBLEM STATEMENT AND CONSTRUCTION OF A RENORMALIZED PERTURBATION THEORY

The goal of the present work is to investigate asymptotics of solutions of the Cauchy problem for a localized initial distribution function. The solution is assumed to be self-similar, i.e., the functional form of the solution does not change in the process of evolution, and only the characteristic scales of the problem, namely, the amplitude and width of the distribution, depend on time. In the case of spherically symmetrical initial conditions, the self-similarity property implies

$$C(\mathbf{r},t)=C(t)F(r^2/l^2(t)), \tag{2.1}$$

and one can describe the evolution of the distribution in terms of the functions $C(t)$ and $l^2(t)$. [It is worth noting that previous studies[7,8] of this problem were limited to the case of $l^2(t)=4D_0t$.] The author previously[9] suggested using the RG method in order to calculate these functions and analyzed the specific case of an absorption reaction in a low-dimensional space. In the present paper this approach is generalized to the case of a space of arbitrary dimension and arbitrary sign of λ, which allows one to investigate, using the same technique, processes involving reactions with material generation and so-called blow-up regimes, which commonly occur in such reactions.[8]

In what follows, we will use the total amount of material $q(t)$ and time-dependent diffusion coefficient $\tilde{D}$ defined by the formula $l^2(t)=4\tilde{D}(t)t$ instead of the functions $C(t)$ and $l^2(t)$. These two parameters can be expressed in terms of $C(t)$ and $l^2(t)$ and allow us to write the solution in the form

$$C(\mathbf{r},t)=\frac{q(t)}{[4\tilde{D}(t)t]^{d/2}b_1}F\left(\frac{r^2}{4\tilde{D}(t)t}\right), \quad b_1=\int d\mathbf{x}\,F(x^2), \tag{2.2}$$

where

$$\int d\mathbf{r}\,C(\mathbf{r},t)=q(t),$$

$$\int d\mathbf{r}\,r^2C(\mathbf{r},t)=4q(t)\tilde{D}(t)t\left(\frac{b_1}{b_2}\right),$$

$$b_2=\int d\mathbf{x}\,x^2F(x^2). \tag{2.3}$$

The numerical analysis of the solution[10] indicates that, for a large set of different initial distributions and at least in the case of an absorption reaction ($\lambda>0$), an asymptotic solution is well described by the Gaussian form $F(x^2)=\exp\{-x^2\}$. In this case we have $b_1=\pi^{d/2}$ and $b_2/b_1=d/2$.

Let us take the initial distribution in the form

$$C(\mathbf{r},0)=Q_0\delta(\mathbf{r}). \tag{2.4}$$

However, in some cases, using a generalized function (distribution) as an initial condition can yield an incorrect result, since the function $\delta^n(\mathbf{r})$ is ill-defined, and then one should use a regularized distribution of the form

$$C(\mathbf{r},0)=C_0\exp(-r^2/l_0^2), \tag{2.4$'$}$$

which can be treated as resulting from linear evolution of the distribution (2.4) defined at $t=-t_0$, so that $t_0=l_0^2/4D_0$ and $Q_0=(\pi l_0^2)^{d/2}C_0$.

The use the RG method in its quantum-field version is based on the arbotrariness of a partition of the full Hamiltonian of a system into unperturbed and perturbed parts when a renormalized perturbation scheme is developed (renormalization invariance), and this arbitrariness is used to improve the perturbation theory by summing an infinite subset of the total perturbation series.[11] Although our version of the theory does not contain a Hamiltonian, and the perturbation theory is developed by iterating an integral equation equivalent to Eqs. (1.1) and (2.4), the selection of the initial distribution provides a degree of freedom which corresponds to the renormalization invariance in the quantum field theory.

Following the RG method, we renormalize the diffusion coefficient in Eq. (1.1) by replacing $D_0\to D=Z_1^{-1}D_0$ and adding a compensating term to the right-hand side:

$$\left[\frac{\partial}{\partial t}-D\Delta\right]C(\mathbf{r},t)+\lambda C^{1+2\delta}(\mathbf{r},t)=(Z_1-1)D\Delta C(\mathbf{r},t). \tag{2.5}$$

Using the Green's function method, let us go over from a differential to an integral equation, incorporating the initial condition (2.4) in the explicit form

$$C(\mathbf{r},t)=Q_0G(\mathbf{r},t)-\int_0^t dt'\int d\mathbf{r}'\,G(\mathbf{r}-\mathbf{r}',t-t') \times[\lambda C^{1+2\delta}(\mathbf{r}',t')-(Z_1-1)D\Delta C(\mathbf{r}',t')] \tag{2.6}$$

[note that the Green's function in Eq. (2.6) incorporates the renormalized diffusion coefficient D rather then its bare value D_0].

Let us renormalize the initial condition by substituting $Q_0\to Q=Z_2^{-1}Q_0$ and adding a second compensating term to the right-hand side:

$$C(\mathbf{r},t)=QG(\mathbf{r},t)-\int_0^t dt'\int d\mathbf{r}'\,G(\mathbf{r}-\mathbf{r}',t-t') \times[\lambda C^{1+2\delta}(\mathbf{r}',t')-(Z_1-1)D\Delta C(\mathbf{r}',t')] +(Z_2-1)QG(\mathbf{r},t). \tag{2.7}$$

After sequential iterations of Eq. (2.7) with the first term in the right-hand side used as a zero-order approximation, we obtain the solution in the form of a perturbation series in powers of the nonlinearity parameter λ.

In the first approximation of the perturbation theory, the calculation yields[9]

$$C^{(1)}(\mathbf{r},t)=Q\Bigg[G(\mathbf{r},t)-\frac{\lambda Q^{2\delta}}{D^{\delta d}}A\int_0^t\frac{dt'}{(t')^{1-\epsilon}}G\times\left(\mathbf{r},t-\frac{2\delta}{1+2\delta}t'\right)+(Z_2-1)G(\mathbf{r},t)+(Z_1-1)Dt\Delta G(\mathbf{r},t)\Bigg],$$

$$\epsilon=1-\delta d,\quad A=\frac{1}{(1+2\delta)^{d/2}}\frac{1}{(4\pi)^{1-\epsilon}}. \tag{2.8}$$

In getting Eq. (2.8), we have used the so-called semigroup property of the Green's function of the diffusion equation, namely,

$$\int d\mathbf{r}'G(\mathbf{r}-\mathbf{r}',t-t')G(\mathbf{r}',t'-t_0)=G(\mathbf{r},t-t_0)\quad(t>t'>t_0), \tag{2.9}$$

and it is easy to show that

$$G^{1+2\delta}(\mathbf{r},t)=\frac{A}{(Dt)^{\delta d}}G\left(\mathbf{r},\frac{t}{1+2\delta}\right). \tag{2.10}$$

Equation (2.8) clearly shows that, at least at $\epsilon<0$, the main contribution to the integral over t' comes from the region of small t', which allows us to omit t' in the argument of the Green's function in the integrand. As a result, we see that in the lowest order of the perturbation theory in $1/\epsilon$ the spatial distribution of the concentration is described by the Green's function of the linear problem. This confirms our assumption (2.1) about the self-similar form of the distribution function, because the same arguments are valid for higher order terms of the perturbation theory constructed by iteration. Note also that in the case of small δ, in accordance with Eq. (2.8), one can also omit t' in the argument of the Green's function and obtain the spatial distribution typical of the linear problem.

Up to this point, we stress that the renormalization constants Z_1 and Z_2 have been arbitrary. We will specify these constants with the normalization condition, which says that the amount of material and the effective diffusion coefficient at time $t=\tau$ coincide with the renormalized values

$$q(\tau)=Q,\quad \tilde{D}(\tau)=D. \tag{2.11}$$

The use of Eqs. (2.3), (2.8), and (2.11) yields in the lowest approximation of the perturbation theory the following renormalization constant for the total amount of material:

$$Z_2(\tau)=1+\frac{\lambda Q^{2\delta}}{D^{\delta d}}A\int_0^\tau\frac{dt'}{(t')^{1-\epsilon}}=1+gAB(1,\epsilon), \tag{2.12}$$

where $g=(\lambda Q^{2\delta}/D^{\delta d})\tau^{\epsilon}$, $B(\xi,\eta)$ is the beta-function (the Euler integral of the first kind). The beta-function is included in Eq. (2.12) because the integral diverges at negative ϵ, but it can be extrapolated to the region $\epsilon<0$ using well-known recurrence relations. This approach corresponds to the method of dimensional regularization in quantum field theory.[12,13]

Similarly we determine the renormalization constant for the diffusion coefficient:

$$Z_1(\tau)=1-\frac{\lambda Q^{2\delta}}{D^{\delta d}}A\frac{2\delta}{1+2\delta}\frac{1}{\tau}\int_0^\tau\frac{dt'}{(t')^{-\epsilon}}=1-2gAaB(1,\epsilon),\quad a=\frac{\delta}{1+2\delta}\frac{B(1,1+\epsilon)}{B(1,\epsilon)}. \tag{2.13}$$

3. RENORMALIZATION INVARIANCE AND RENORMALIZATION GROUP METHOD

Let us express the unknown functions $q(t)$ and $\tilde{D}(t)$ in terms of dimensionless functions of dimensionless parameters:

$$q(t)=Q\,f_2(t/\tau,g),\quad \tilde{D}(t)=D\,f_1(t/\tau,g), \tag{3.1}$$

which should obey the relations $f_1(1,g)=1$ and $f_2(1,g)=1$ according to the normalization condition (2.11). The renormalization invariance requirement means that calculations of $q(t)$ and $\tilde{D}(t)$ should not depend on the set of numerical parameters determined by a specific selection of normalization point τ, i.e., under the RG transformations $\tau\to\tau_1$, $Q\to Q_1$, and $D\to D_1$ the following relation should hold:

$$Qf_2\left(\frac{t}{\tau},g\right)=Q_1f_2\left(\frac{t}{\tau_1},g_1\right),\quad Df_1\left(\frac{t}{\tau},g\right)=D_1f_1\left(\frac{t}{\tau_1},g_1\right),$$

$$g_1=\frac{\lambda Q_1^{2\delta}}{D_1^{\delta d}}\tau_1^{\epsilon}. \tag{3.2}$$

From the normalization conditions it follows that

$$Q_1=Qf_2\left(\frac{\tau_1}{\tau},g\right),\quad D_1=Df_1\left(\frac{\tau_1}{\tau},g\right),$$

which yields functional RG equations

$$f_i\left(\frac{t}{\tau},g\right)=f_i\left(\frac{\tau_1}{\tau},g\right)f_i\left(\frac{t}{\tau_1},g_1\right)\quad(i=1,2). \tag{3.3}$$

To solve Eq. (3.3), let us introduce a new dimensionless function

$$\frac{\lambda\,q^{2\delta}(t)t^{\epsilon}}{\tilde{D}^{\delta d}(t)}\equiv g\,f_2^{2\delta}\left(\frac{t}{\tau},g\right)\left(\frac{t}{\tau}\right)^{\epsilon}f_1^{\delta d}\left(\frac{t}{\tau},g\right)=\tilde{g}\left(\frac{t}{\tau},g\right). \tag{3.4}$$

This function is an actual time-dependent expansion parameter of the perturbation theory invariant under the RG transformation. It satisfies the functional RG equation

$$\tilde{g}(x,g)=\tilde{g}(x/\alpha,\tilde{g}(\alpha,g)),\quad \tilde{g}(1,g)=g \tag{3.5}$$

and the differential RG equation following from it:

$$\left\{-x\frac{\partial}{\partial x}+\beta(g)\frac{\partial}{\partial g}\right\}\tilde{g}(x,g)=0,$$

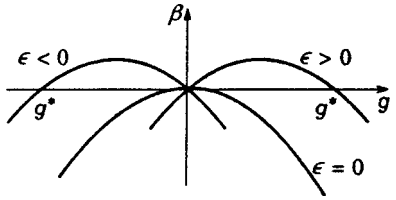

FIG. 1. Possible forms of $\beta(g)$.

$$\beta(g)=\left.\frac{\partial\tilde{g}(x,g)}{\partial x}\right|_{x=1}. \tag{3.6}$$

In reality, Eq. (3.6) can be treated as an equation only in the case of a given RG function $\beta(g)$, which is determined by the behavior of function $\tilde{g}(x,g)$ near the normalization point $x=1$. The RG method is based on the suggestion that $\beta(g)$ should be calculated by the renormalized perturbation theory. If β is calculated by the perturbation theory, substituted in Eq. (3.6), and this equation is solved, this procedure corresponds to summation of the perturbation series or an infinite subsequence of it.[11]

Using Eqs. (2.12) and (2.13), one can evaluate the functions $f_i(x,g)$ in the lowest order of perturbation theory:

$$\begin{aligned} f_1(x,g)&=\tilde{D}(t)/D=Z_1(\tau)/Z_1(t)\\ &\approx 1+2gAaB(1,\epsilon)(x^\epsilon-1),\\ f_2(x,g)&=q(t)/Q=Z_2(\tau)/Z_2(t)\approx 1-gAB(1,\epsilon)(x^\epsilon-1). \end{aligned} \tag{3.7}$$

After finding $\beta(g)$ from Eq. (3.7) and solving Eq. (3.6) by the method of characteristics,[11] we get

$$\begin{aligned} &\tilde{g}(x,g)=\frac{gx^\epsilon}{1+(g/g^*)(x^\epsilon-1)},\\ &f_i(x,g)=\{1+(g/g^*)(x^\epsilon-1)\}^{\zeta_i},\\ &\zeta_1=a/\delta(1+ad),\quad \zeta_2=-1/2\delta(1+ad),\\ &g^*=\epsilon/2\delta A(1+ad). \end{aligned} \tag{3.8}$$

The remaining step is to express the solution in terms of initial (nonrenormalized) parameters of the problem by eliminating τ, D, and Q. To this end, let us consider the general form of the RG function. Depending on the sign of ϵ, three situations illustrated by Fig. 1 can occur. The asymptotics is determined by positions of stationary (fixed) points g_i determined by the condition $\beta(g_i)=0$. The requirement that the fixed point should be stable as $t\to\infty$ is equivalent to the condition $[\partial\beta(g)/\partial g]_{g=g_i}<0$, and the stability condition for a fixed point as $t\to 0$ is $[\partial\beta(g)/\partial g]_{g=g_i}>0$. Figure 1 shows that in the lowest approximation of perturbation theory there are two fixed points: one is trivial, $g=0$, which corresponds to the absence of nonlinear interactions (asymptotic freedom), and the other is nontrivial, $g=g^*\neq 0$, whose position is determined by the sign of ϵ. In the marginal case $\epsilon=0$ the stable and unstable fixed points are merged.

From the condition $q(t)\geqslant 0$, $\tilde{D}(t)\geqslant 0$ it follows that the sign of g is determined by the sign of λ only. Hence it follows that in the case of an absorption reaction ($\lambda>0$) and $\epsilon>0$ ($d<d_c$) a specific stable nonlinear regime develops as $t\to\infty$ and it evolves from the initial condition of asymptotic freedom as $t\to 0$. For $\epsilon<0$ ($d>d_c$), an asymptotically free regime may occur as $t\to\infty$, but it evolves from ''unphysical'' boundary conditions defined at $t\to 0$ and corresponding to $g<0$.

For $\lambda<0$ (the case when a substance is produced) and $\epsilon>0$, a regime stable as $t\to\infty$ corresponds to the unphysical value $\tilde{g}\to g^*>0$, so the region $t\to\infty$ should be excluded from the analysis. But for $\epsilon<0$ the regime of asymptotic freedom in the limit $t\to\infty$ evolves from an initial state that is stable for $t\to 0$ and corresponding to $g<0$.

The condition of asymptotic freedom means an absence of renormalization of the diffusion coefficient and a constant amount of material. These requirements can be treated as boundary conditions that allow one to eliminate τ in Eq. (3.8) and express Q and D in terms of the initial parameters. In accordance with the above arguments, this procedure is carried out in a different manner depending on the signs of λ and ϵ.

4. DIFFERENT REGIMES OF SYSTEM BEHAVIOR

(a) An absorption reaction at space dimension below the critical value, $\lambda>0$, $\epsilon>0$ (binary reactions in a one-dimensional space). Using the condition of asymptotic freedom, $q(0)=Q_0$, $\tilde{D}(0)=D_0$, we get from Eqs. (3.4) and (3.8)

$$g=g_0[1+(g_0/g^*)]^{-1},\quad g_0=\lambda Q_0^{2\delta}\tau^\epsilon/D_0^{\delta d}, \tag{4.1}$$

which yields after substitution in Eq. (3.1)

$$\begin{aligned} &\tilde{D}(t)=D_0\left[1+\frac{\lambda Q_0^{2\delta}}{D_0^{\delta d}}\frac{1}{g^*}t^\epsilon\right]^{\zeta_1}=D_0\left[1+\frac{1}{g^*}\left(\frac{t}{T}\right)^\epsilon\right]^{\zeta_1},\\ &q(t)=Q_0\left[1+\frac{\lambda Q_0^{2\delta}}{D_0^{\delta d}}\frac{1}{g^*}t^\epsilon\right]^{\zeta_2}=Q_0\left[1+\frac{1}{g^*}\left(\frac{t}{T}\right)^\epsilon\right]^{\zeta_2},\\ &T^{-\epsilon}=\frac{\lambda Q_0^{2\delta}}{D_0^{\delta d}}. \end{aligned} \tag{4.2}$$

From Eq. (4.2) we obtain expressions for the amplitude $C(t)$ and radius $l(t)$ of the diffusion spot:

$$\begin{aligned} C(t)&=\frac{q(t)}{[4\pi\tilde{D}(t)t]^{d/2}}\\ &=\frac{Q_0}{[4\pi D_0 t]^{d/2}}\left[1+\frac{1}{g^*}\left(\frac{t}{T}\right)^\epsilon\right]^{-1/2\delta},\\ l(t)&=\sqrt{4\tilde{D}(t)t}=\sqrt{4D_0}\left[1+\frac{1}{g^*}\left(\frac{t}{T}\right)^\epsilon\right]^{\zeta_1/2}t^{1/2}. \end{aligned} \tag{4.3}$$

At large t Eqs. (4.3) transforms to

$$\begin{aligned} &C(t)\to\frac{1}{(4\pi)^{d/2}}\left(\frac{\lambda t}{g^*}\right)^{-1/2\delta}=c(\lambda t)^{-1/(n-1)},\\ &l(t)\to\sqrt{4D_0}(g^*)^{-\zeta_2/2}t^{1/2}(t/T)^{\epsilon\zeta_1/2}. \end{aligned} \tag{4.4}$$

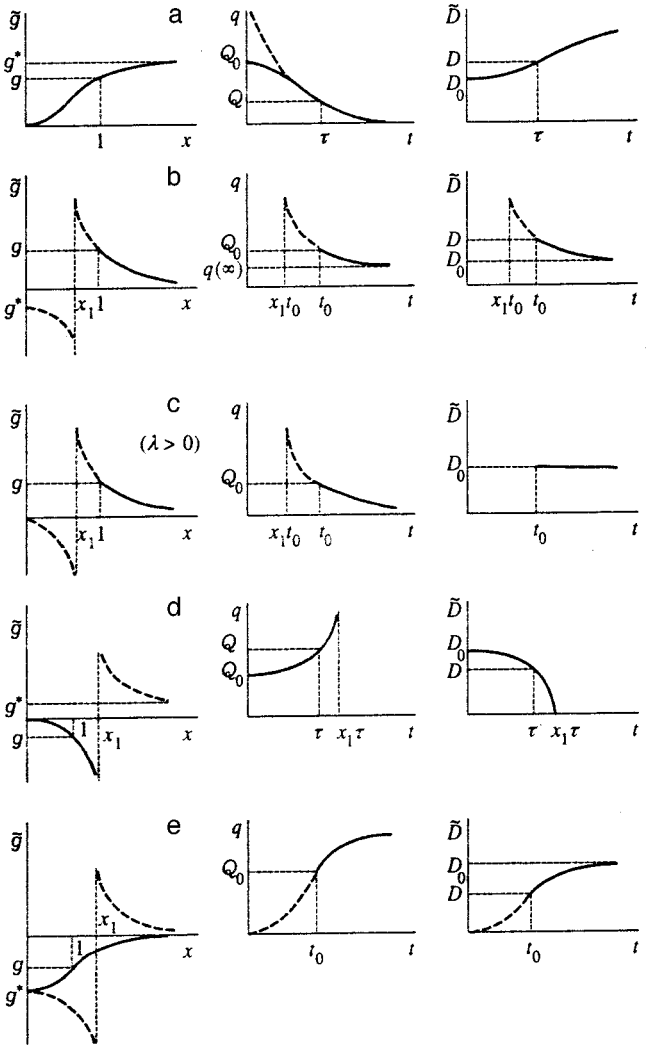

FIG. 2. Typical curves of $\tilde{g}(x,g)$, $q(t)$, and $\tilde{D}(t)$ for various ϵ and λ [see text, cases (a) to (e)].

In accordance with Eq. (4.4), the asymptotic form of the amplitude proves to be independent of the initial distribution parameter Q_0 and corresponds to self-similar solutions discussed in earlier publications.[7] However, the evolution of the diffusion spot radius $l(t)$ depends on the selection of initial conditions (parameter Q_0), i.e., a regime of incomplete self-similar condition is realized (self-similarity of the second kind),[14] when $l(t) \propto t^{1/2+\alpha}$. The incomplete self-similarity exponent α is determined only by the reaction order and spatial dimensionality:

$$\alpha = \frac{1-\delta d}{2\delta}\frac{a}{1+ad}.$$

The corresponding curves are plotted in Fig. 2a.

(b) Absorption reaction in a space of dimension above the critical value, $\epsilon<0$, $\lambda>0$. The asymptotic freedom regime realized as $t\to\infty$ corresponds to the conditions $\tilde{D}(\infty)=D_0$ and $q(\infty)=\text{const}$. However, in the limit $t\to 0$, it is found that $\tilde{g}\to g^*<0$, which corresponds to unphysical initial conditions since, according to Eq. (3.3), the relation $\tilde{g}\geqslant 0$ should hold. The reason is that the initial condition is defined by the generalized function $\delta(\mathbf{r})$: the function $\delta^n(\mathbf{r})$ appearing in Eq. (1.1) is ill-defined. In order to get circumvent this difficulty, one should take an initial condition with a finite distribution radius [Eq. (2.4′)] $l_0=\sqrt{4\tilde{D}(t_0)t_0}$ and t_0 in the region of positive $\tilde{g}$. Thus, the initial conditions are given by the relations

$$Q_0 = q(t_0), \quad l_0 = l(t_0). \tag{4.5}$$

Having selected the normalization point at $\tau=t_0$ and taken into account $\tilde{D}(t_0)=D$, we obtain

$$q(t) = Q_0\left[1+\frac{\lambda Q_0^{2\delta}}{D^{\delta d}}\frac{1}{(-g^*)}(t_0^{\epsilon}-t^{\epsilon})\right]^{\zeta_2},$$

$$\tilde{D}(t) = D\left[1+\frac{\lambda Q_0^{2\delta}}{D^{\delta d}}\frac{1}{(-g^*)}(t_0^{\epsilon}-t^{\epsilon})\right]^{\zeta_1}. \tag{4.6}$$

From condition $\tilde{D}(\infty)=D_0$ we derive an equation expressing t_0 in terms of the initial conditions, $C_0=C(t_0)$ and l_0, and the parameters λ and D_0 of the problem:

$$4D_0t_0 = l_0^2\left[1+\frac{(4\pi)^{\delta d}}{(-g^*)}C_0^{2\delta}\lambda t_0\right]^{\zeta_1}. \tag{4.7}$$

For the amplitude and radius of the distribution given by Eq. (2.1), we find

$$C(t) = C_0\left(\frac{t}{t_0}\right)^{-d/2}\left\{1+\frac{(4\pi)^{\delta d}}{(-g^*)}C_0^{2\delta}\lambda t_0 \times\left[1-\left(\frac{t}{t_0}\right)^{\epsilon}\right]\right\}^{-1/2\delta},$$

$$l(t) = l_0\left(\frac{t}{t_0}\right)^{1/2}\left\{1+\frac{(4\pi)^{\delta d}}{(-g^*)}C_0^{2\delta}\lambda t_0\left[1-\left(\frac{t}{t_0}\right)^{\epsilon}\right]\right\}^{\zeta_1/2}, \tag{4.8}$$

and for the asymptotic amount of material

$$q(\infty) = Q_0(4D_0t_0/l_0^2)^{-1/2a}. \tag{4.9}$$

Typical curves for $\tilde{g}(x)$, $q(t)$, and $\tilde{D}(t)$ are shown in Fig. 2b.

(c) The marginal case of the critical dimension, $\epsilon=0$ (takes place in binary reactions in a two-dimensional space). An important point is that at $\epsilon=0$ we have $\zeta_1=0$, in accordance with Eq. (2.13), and the diffusion coefficient is not renormalized. Then, using the formula for g^* from Eq. (3.8), we obtain

$$q(t) = Q_0\{1+2\delta A g\ln(t/\tau)\}^{-1/2\delta}, \quad \tilde{D}(t)=D_0. \tag{4.10}$$

By setting the initial conditions at the renormalization point $\tau=l_0^2/4D_0$ and measuring time from this point we obtain

$$q(t) = Q_0\left[1+\frac{1}{\pi(d+2)}\frac{\lambda Q_0^{1/2d}}{D_0}\ln\frac{l(t)}{l_0}\right]^{-d/2},$$

$$l(t) = \sqrt{l_0^2+4D_0t}. \tag{4.11}$$

From Eq. (4.11) it follows that in the case of an absorption reaction ($\lambda>0$) the amount of material in the long-time limit

slowly decays as a power of a logarithm and the asymptotic form is independent of the initial quantity Q_0. However, in the case of the production reaction ($\lambda<0$) the amount of material increases and tends to infinity during a certain finite time interval depending on the problem parameters λ, D_0, Q_0, and l_0.

(d) A creation reaction at a dimension below the critical value, $\epsilon>0$, $\lambda<0$ (a binary reaction in a one-dimensional space). After eliminating the renormalized parameter with the help of the condition of asymptotic freedom for $t\to 0$, we obtain formulas of the form [Eq. (4.2)]

$$\tilde{D}(t)=D_0\left[1-\frac{(-\lambda)Q_0^{2\delta}}{D_0^{\delta d}}\frac{1}{g^*}t^{\epsilon}\right]^{\zeta_1},$$

$$q(t)=Q_0\left[1-\frac{(-\lambda)Q_0^{2\delta}}{D_0^{\delta d}}\frac{1}{g^*}t^{\epsilon}\right]^{\zeta_2}. \tag{4.12}$$

With due account of the conditions $\zeta_1>0$ and $\zeta_2<0$, we find that $q(t)$ grows with time and has a singularity at $t_1=[(-\lambda)Q_0^{2\delta}/g^*D_0^{\delta d}]^{-1/\epsilon}$ (Fig. 2d). It also follows from Eq. (4.12) that at first an initially localized distribution of a finite amount of material spreads in accordance with the linear theory, but then the distribution narrows as the amount of material increases further, and a collapse takes place in a finite time t_1, which corresponds to the peaking regime.[8] An investigation of the behavior of the system near the singularity point probably requires that the functions $f_i(x,g)$ be calculated in higher orders of the perturbation theory, but the present analysis probably allows one to describe the behavior and evolution rates of parameters in the peaking regime, at least in its earlier stages.

(e) A reaction producing material at a dimension above the critical value, $\epsilon<0$, $\lambda<0$. It follows from the general equations (3.8) with due account of conditions $\zeta_1<0$ and $\zeta_2<0$ that the solutions of RG equations for $q(t)$ and $\tilde{D}(t)$ tend to zero as $t\to 0$, which means that the Cauchy problem with the initial δ-function distribution is ill-posed. Therefore, one should proceed as in case (b), when the initial time was chosen at $t=t_0$ and the normalization conditions were determined at this point. Analysis of the solutions indicates that in this case two different regimes are possible, corresponding to the situations illustrated by the solid ($g/g^*<1$) and dashed ($g/g^*>1$) curves in Fig. 2e. It is clear that for $g/g^*>1$ the region $x>x_1$ is unphysical since $\tilde{g}>0$ at $\lambda<0$.

In accordance with the normalization conditions, Eq. (4.5) and the condition of asymptotic freedom $\tilde{D}(\infty)=D_0$ must be treated as boundary conditions and used in expressing integration constants Q, D, and $\tau=t_0$ of RG equations, in terms of the initial parameters that determine the total amount of material Q_0 and distribution width l_0, or the initial amplitude $C_0=Q_0/\pi^{d/2}l_0^d$ and l_0. The condition of asymptotic freedom can be used only at $g/g^*<1$ since the region of asymptotic behavior is unphysical at $g/g^*>1$ and the problem is ill-posed.

By eliminating the intermediate parameters g, Q, and D we obtain equations of the form (4.6) and (4.7), and the existence condition for the solution t_0 is $g/g^*<1$. If the initial conditions are chosen so that $g=g^*$, the solutions have the form

$$\tilde{g}(x,g)=g^*,\quad q(t)=Q_0\left(\frac{t}{t_0}\right)^{\epsilon\zeta_2},\quad l(t)=l_0\left(\frac{t}{t_0}\right)^{\epsilon\zeta_1+1/2},$$

$$C(t)=\left[\frac{1+2\delta}{2\delta}\frac{\delta d-1}{1+ad}\right]^{-1/2\delta}(-\lambda t)^{-1/2\delta}\equiv c(-\lambda t)^{-1/2\delta}. \tag{4.13}$$

Thus, the amplitude in the degenerate case under consideration is independent of the initial conditions, whereas the amount of material $q(t)$ and diffusion spot diameter $l(t)$ contain functions of time that correspond to the incomplete self-similarity condition. It follows from the last line in Eq. (4.13) that t_0 can be derived from the initial amplitude C_0 and, after substituting in the expression for $q(t)$ and $l(t)$, one can calculate the numerical coefficients in the functions of time.

5. DISCUSSION

The paper suggests a new version of the RG technique applied to chemical reactions controlled by diffusion. Unlike the traditional RG technique, which renormalizes the characteristic parameters of a system such as masses and charges in field theory or similar parameters in other problems, in this paper the initial conditions, namely, the amplitude and radius of the distribution function, have been renormalized. In this connection, the renormalization invariance manifests itself in the explicit form as a system behavior independence of given initial (boundary) conditions, which is dubbed the functional self-similarity.[15] Although in most cases the functional self-similarity leads to trivial results in the sense that RG differential equations prove to be identical to initial equations, in the example discussed in the paper the RG technique yields nontrivial solutions. Previously the functional self-similarity property was used in solving the problem of nonlinear diffusion when asymptotic solutions of equations of a special type were investigated.[16,17]

The method used in this work has yielded results that are sometimes different from previously known solutions. In particular, in the case of an absorption reaction, the asymptotic amplitude decays according to a power law with exponent $-1/(n-1)$ [Eq. (4.4)] for spatial dimensions below the critical value, whereas, according to the conventional viewpoint, this law should apply only to dimensions above the critical value.[3–5] The reason for this disagreement is that earlier workers operated assumed the mean-field approximation, according to which field inhomogeneities due to fluctuations are small and the dominant role is played by nonlinear damping processes. But in this paper we have analyzed the evolution of an initially localized distribution instead of fluctuations in a spatially homogeneous distribution, where the distribution does not spread with time and the distribution amplitude follows the same law as the total amount of material.

The existence of self-similar solutions of the form (4.4) for the amplitude was proven on the basis of different arguments[7] when self-similar asymptotic forms of solutions

of a quasilinear parabolic equation independent of initial conditions (''eigenfunctions'') were found. On these assumptions, the amplitude decay exponent followed from the dimensions of physical quantities $[-1/(n-1)]$, and it was proven that a distribution decaying faster than a power law as time tends to infinity can exist only if the spatial dimension is below the critical value [case (a)]. On the contrary, if the dimension is higher than the critical one and the nonlinearity degree is high, the amplitude decays rapidly on the initial stage, and as a result, the nonlinear term is unimportant and the system evolves in the regime of asymptotic freedom with diffusive spreading in accordance with the linear theory and conservation of the total amount of material. It turns out that the regime of asymptotic freedom cannot evolve from any initial distribution, but only under certain limitations on the amplitude and width. This limitation is defined by the domain of real solutions of Eq. (4.7) for parameter t_0. Logarithmic functions of time arise in the RG method in a natural way at the critical dimension.[5] This result was discovered previously using different techniques (see the review[7]).

Using the RG technique, we have also described in a natural manner peaking regimes in the case of a reaction generating material at dimensions below the critical one, which were discovered and described previously in analyzing thermal processes by means of numerical and semianalytical methods.[8] Application of the RG method to the problem under discussion allows one to estimate convergence rates and times of distribution collapse as functions of problem parameters.

As concerns the problems of absorption reactions, in case (a) the traditional ϵ-expansion yields results of the physical dimension analysis, provided that the asymptotic form is independent of initial conditions (universal). However, our analysis indicates that only the amplitude is described by a universal function. Nonetheless, the universality is incomplete, because an asymptotic growth of the distribution width corresponds to the presence of anomalous dimension, and the additional dimensional parameter due to initial conditions does not disappear in the long-time limit. The RG method has not yet been applied to problems with material generation.

The analysis of solutions of RG equations not only makes it possible to get the exponents of power functions in the asymptotic long-time limit, it also allows us to track the transition to the asymptotic regime and calculate the numerical coefficients.[6,9]

The work was supported by the Russian Fund for Fundamental Research (Grant 96-01-00748).

*)E-mail: teodor@ipmnet.ru

[1] D. A. Frank-Kamenetskii, *Diffusion and Heat Transfer in Chemical Kinetics*, 2nd, ed., Plenum, New York (1969) [Russian original, Nauka, Moscow (1987)].
[2] P. C. Martin, E. D. Siggia, and H. A. Rose, Phys. Rev. A **8**, 423 (1973).
[3] L. Peliti, J. Phys. A **19**, L365 (1986).
[4] M. Doi, J. Phys. A **9**, 1465 (1976).
[5] T. Ohtsuki, Phys. Rev. E **43**, 6917 (1991).
[6] B. P. Lee, J. Phys. A **27**, 2633 (1994).
[7] V. A. Galaktionov, S. P. Kurdyumov, and A. A. Samarskii, Matematicheskii Sbornik **126**, 435 (1986).
[8] A. A. Samarskii, V. A. Galaktionov, S. P. Kurdyumov, and A. P. Mikhailov, *Blow-up in Quasilinear Parabolic Equations*, W. de Gruyter, Berlin, New York (1995) [Russian original, Nauka, Moscow (1987)].
[9] É. V. Teodorovich, Prikl. Mat. Mekh. **62**, 443 (1998).
[10] V. M. Entov (private communication).
[11] N. N. Bogolyubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, 3rd ed., Wiley, New York (1980) [Russian original, Nauka, Moscow (1984)].
[12] P. Ramond, *Field Theory*, Benjamin/Cummings, Reading (USA) (1981).
[13] J. Collins, *Renormalization*, Cambridge University Press (1984).
[14] G. I. Barenblatt, *Scaling, Self-Similarity, and Intermediate Asymptotic*, Cambridge University Press, Cambridge (1996) [Russian original, Gidrometeoizdat, Leningrad (1978)].
[15] D. V. Shirkov, Dokl. Akad. Nauk SSSR **263**, 64 (1982) [Sov. Phys. Dokl. **27**, 197 (1982)].
[16] N. Goldenfeld, O. Martin, Y. Oono, and Fong Liu, Phys. Rev. Lett. **64**, 1361 (1990).
[17] I. S. Ginzburg, V. M. Entov, and É. V. Teodorovich, Prikl. Mat. Mekh. **52**, 68 (1992); V. M. Entov and É. V. Teodorovich, Teor. Osnovy Khim. Tekhnologii, No. 3, 218 (1993).

Translation provided by the Russian Editorial office.

Dynamic topological solitons in a two-dimensional ferromagnet

A. A. Zhmudskiĭ and B. A. Ivanov*)

Institute of Magnetism, Ukrainian National Academy of Sciences, 252680 Kiev, Ukraine
(Submitted 19 August 1998)
Zh. Éksp. Teor. Fiz. **115**, 1511–1530 (April 1999)

It is shown that stable, skyrmion-type, dynamic solitons can be constructed for a wide class of two-dimensional models of anisotropic ferromagnets. These solitons are stabilized as a result of the conservation of various integrals of motion: the z projection of the total spin S_z or the orbital angular momentum L_z of the magnetization field. A class of two-parameter solitons with quite complicated (almost periodic) magnetization-field dynamics exists for a purely uniaxial model (in the sense of both spin and spatial rotations) with maximum symmetry. Stable solitons with periodic magnetization dynamics exist for ferromagnets with lower symmetry (only S_z or L_z or the total angular momentum $J_z=L_z+S_z$ is conserved). © *1999 American Institute of Physics.* [S1063-7761(99)02504-4]

1. INTRODUCTION

Nonlinear excitations—topological magnetic solitons (see Ref. 1)—play an important role in the physics of low-dimensional magnets[2,3] Specifically, it is well known that magnetic vortices contribute to the physical properties of two-dimensional magnets with continuous degeneracy of the ground state (Refs. 4–6; see also Refs. 2 and 3). Vortex solutions are impossible in two-dimensional magnets with discrete degeneracy of the ground state, i.e., in magnets with easy-axis anisotropy (or orthorhombic magnets). For these magnets it is important to take into account stable (quite long-lived), localized, two-dimensional solitons.[2,3] According to experiments,[7,8] they determine the relaxation of magnetic excitations and can give rise to peaks in the response functions.[9] Such topological (nontrivial homotopy group π_2) statistical solitons have been constructed by Belavin and Polyakov for two-dimensional isotropic magnets.[10] They are characterized by nontrivial topological properties relative to the mapping of the xy plane of a two-dimensional magnet onto the sphere S^2: $\mathbf{m}^2=1$, where $\mathbf{m}$ is the normalized magnetization of the ferromagnet (see Refs. 1–3. In recent years interest in such states has increased even more in connection with their application for describing the Hall quantum effect.[11]

The basic problem arising in the soliton physics of two-dimensional magnets is related to the stability of the localized solitons. According to the Hobart–Derrick theorem,[12] for models of nonlinear fields whose energy depends on the components of the field and is quadratic in the gradients of the field components (see Eq. (1) below) stable, static, non-one-dimensional solitons with finite energy and finite radius do not exist (the solitons are unstable against collapse). Specifically, this is true for a uniaxial two-dimensional ferromagnet characterized by an energy of the form

$$W=\frac{1}{2}\int d^2x\{A(\nabla\cdot\mathbf{m})^2+K(m_x^2+m_y^2)\}, \tag{1}$$

where A is the exchange interaction constant, K is the anisotropy constant, and the z axis is chosen along the easy magnetization axis of the ferromagnet. Exact Belavin–Polyakov solutions[10] for the model (1) exist only in the isotropic state ($K=0$). We note that the model (1) with anisotropy ($K\neq 0$) is fundamentally different from the isotropic case ($K=0$). In the isotropic model the problem is scale-invariant. It is described by a self-duality equation and is exactly integrable in the static case.[10]

Its simplest solution describes a soliton with topological charge ν. In angular variables for the magnetization vector $[m_z=\cos\theta,\ m_x+im_y=\sin\theta\exp(i\varphi)]$ it has the form[10]

$$\tan\frac{\theta}{2}=\left(\frac{R}{r}\right)^{\nu},\quad \varphi=\nu\chi+\varphi_0, \tag{2}$$

where r and χ are polar coordinates in the plane of the magnet, and R and φ_0 are arbitrary constants. The energy of such a soliton in the exchange approximation is given by the formula

$$E_n^{(0)}=E_0|\nu|,\quad E_0=4\pi A. \tag{3}$$

The fact that the energy is independent of R is a result of the scale invariance of the model (1) with $K=0$. When anisotropy [specifically, of the type in Eq. (1)] and the Zeeman energy $w_H\sim(1-\cos\theta)H$ in an external field $\mathbf{H}=H\mathbf{e}_z$ are taken into account, the expression for the soliton energy acquires a term proportional to R^2. The energy has no minimum for any $R\neq 0$. This signifies the absence of static soliton solutions. Sometimes this fact is referred to as an instability of the soliton against collapse.

However, stable, stationary, dynamic solitons can exist for a number of models (see Refs. 1, 2, and 12). Their existence is due ultimately to the presence of an integral of motion whose value does not vanish in the static limit.[1–3,12] Specifically, for the uniaxial ferromagnet (1) the existence of such solitons could be due to the conservation of the z projection S_z of the total spin[1–3] or the z projection L_z of the orbital angular momentum of the magnetization field.[13,14] Since L_z is negative in a soliton with $\nu>0$ (see Refs. 1 and

2), it is convenient to use the quantity $L=-L_z$. In units of the Planck constant $\hbar$ (below we take $\hbar=1$) the quantities $S=S_z$ and L can be expressed by the formulas (see Refs. 1–3)

$$L=-L_z=\frac{s}{a^2}\int d^2x(1-\cos\theta)[\mathbf{r}\cdot\nabla\varphi]_z,$$

$$S=\frac{s}{a^2}\int d^2x(1-\cos\theta), \tag{4}$$

where s is the atomic spin and a is the lattice constant.

Another natural integral of motion—the momentum **P**—can stabilize three-dimensional solitons with nonzero Hopf index,[15] two-dimensional topological solitons in antiferromagnets,[16] and nontopological two-dimensional skyrmion–antiskyrmion pair solitons in an isotropic[17] and easy-plane ferromagnet. However, the conservation of **P** does not stabilize two-dimensional topological solitons in a ferromagnet. For this reason, we shall consider the stabilization of solitons only as a result of the conservation of the angular momentum (4).

We shall consider the model (1) with the maximum admissable symmetry (both L and S are conserved) as well as a model with a lower symmetry, for example, a model described by an energy of the form

$$W=\int d^2x\left\{\frac{1}{2}A_{ik}\nabla_i\mathbf{m}\cdot\nabla_k\mathbf{m}+W_a(m_x,m_y)\right\}. \tag{1′}$$

Here the form of the exchange constants tensor A_{ik} and of the anisotropy energy W_a are determined by the symmetry of the magnet. Depending on the form of A_{ik} and W_a either S (if $W_a=W_a(\theta)$) or L (if $A_{ik}\propto\delta_{ik}$) can be conserved. We shall also discuss a model with $L\neq$ const and $S\neq$ const but where the z projection $J=L_z+S_z=S-L$ of the total angular momentum is conserved. Analysis showed that for all cases enumerated above with at least one integral of motion L, S, or J stable dynamical solitons do exist, and for the highest-symmetry model (1) solitons with almost periodic magnetization dynamics exist.

2. FORM OF THE DYNAMICAL SOLITON SOLUTIONS FOR VARIOUS MODELS OF FERROMAGNETS

Solitons are determined by localized ($\theta\to 0$, $\nabla\theta\to 0$ as $|\mathbf{r}|\to\infty$) solutions of the Landau–Lifshitz equation. In angular variables for the magnetization vector it has the form of a system of two time-dependent partial differenti al equations for the functions $\theta=\theta(x,y,t)$, $\varphi=\varphi(x,y,t)$. For the model (1′)

$$A_{ik}[\nabla_i\nabla_k\theta-\sin\theta\cos\theta(\nabla_i\varphi)(\nabla_k\varphi)]-\frac{\partial W_a}{\partial\theta}$$
$$=-\frac{M_0}{\gamma}\sin\theta\frac{\partial\varphi}{\partial t},$$
$$A_{ik}\nabla_i(\sin^2\theta\nabla_k\varphi)-\frac{\partial W_a}{\partial\varphi}=\frac{M_0}{\gamma}\sin\theta\frac{\partial\theta}{\partial t}, \tag{5}$$

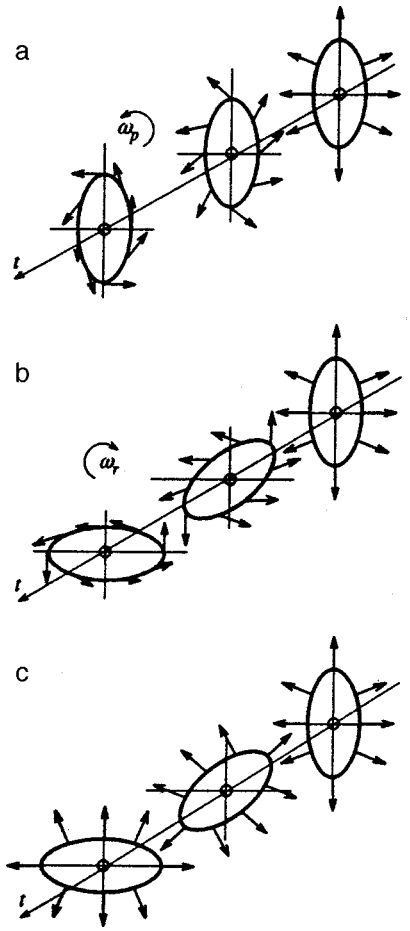

FIG. 1. Magnetization distribution (shown schematically) at the times $t=0$, $t=T/8$, $t=T/4$ ($T=2\pi/\omega$) for solitons in various models: a — precession solitons, $S=$const, $L\neq$ const; b — rotation solitons in the model with $L=$ const, $S\neq$const; c — oscillation solitons in the model with $J=L_z+S_z=$ const; $L\neq$const, $S\neq$const. The arrows show the magnetization directions on the line $\theta=\pi/2$, the symbol + denotes the center of the soliton where $\theta=\pi$.

where $\gamma=2|\mu_B|/\hbar$, μ_B is the Bohr magneton, and M_0 is the saturation magnetization. For the model (1) we have $W_a=K\sin^2\theta$, which does not depend on φ ($A_{ik}\propto\delta_{ik}$).

We shall now discuss the possible form of the dynamical stationary solutions of Eqs. (5). The character of the magnetization distribution in such a soliton and its time dependence can be indicated without solving the dynamical equations for the magnetization, but rather proceeding only from the symmetry of the problem; see Fig. 1. First, we note that the symmetry group of the energy (1) includes rotations of the spins around the z axis and, independently, a rotation of the coordinate axes x and y. The existence of a continuous symmetry results in the appearance of a corresponding integral of motion. Specifically, the possibility of spin rotations exists for all magnets with $W_a=W_a(\theta)$. This symmetry implies conservation of S. If $W_a=W_a(\theta)$, Eqs. (5) do not contain φ and precession solitons can exist. The latter have the form[1–3]

$$\theta=\theta(x,y),\quad \varphi=\omega_p t+\psi(x,y). \tag{6}$$

Here $\theta(x,y)$ and $\varphi(x,y)$ satisfy equations with no time derivatives, as a result of which a soliton of the form (6) is dynamic but stationary. The invariance of the problem under

a rotation of the coordinate axes with $A_{ik}=A\delta_{ik}$ in Eqs. (1′) and (5) implies conservation of the z projection L of the orbital angular momentum and can give rise to rotation solitons (Fig. 1b).[13,14] Such a soliton corresponds to a dynamic but stationary solution in the rotating coordinate system

$$\tilde{x}=x\cos\omega_r t-y\sin\omega_r t,\quad \tilde{y}=x\sin\omega_r t+y\cos\omega_r t \tag{7}$$

of the Landau–Lifshitz equation, of the form

$$\theta=\theta(\tilde{x},\tilde{y}),\quad \varphi=\psi(\tilde{x},\tilde{y}). \tag{8}$$

Only one-parameter solitons of the form (6) or (8) have been discussed previously. However, it is obvious that for the model (1) with the maximum admissable symmetry a two-parameter solution of the form

$$\theta=\theta(\tilde{x},\tilde{y}),\quad \varphi=\omega_p t+\psi(\tilde{x},\tilde{y}), \tag{9}$$

where ω_p and ω_r are two independent parameters, can be considered. In this case for $n\omega_r\neq m\omega_p$, where n and m are integers, the magnetization is an almost periodic function of time.

It is easy to see that with the substitutions (6)–(9) the nonstationary Landau–Lifshitz equation with a definite symmetry of the energy can become stationary (in the rotating coordinate system $\tilde{x}$, $\tilde{y}$). Defining the polar coordinates $\tilde{r}=\sqrt{\tilde{x}^2+\tilde{y}^2}$, $\tilde{\chi}=\tan^{-1}(\tilde{y}/\tilde{x})$ and taking account of the fact that $d\tilde{\chi}/dt=\omega_r$, the right-hand sides of Eqs. (5) can be easily written as

$$\frac{\partial\varphi}{\partial t}=\omega_p+\omega_r\frac{\partial\varphi}{\partial\tilde{\chi}},\quad \frac{\partial\theta}{\partial t}=\omega_r\frac{\partial\theta}{\partial\tilde{\chi}}. \tag{10}$$

Analysis of Eqs. (8)–(10) makes it possible to explain clearly the possibility of using any particular substitution for various models of ferromagnets. Specifically, in the model (1) the energy density of the ferromagnet is independent of the angular variable φ (see Eq. (16) below) and the coordinate $\tilde{\chi}$ (since $A_{ik}\propto\delta_{ik}$). It is obvious that in this case a substitution of the general form (9) yields a stationary equation for the functions $\theta(\tilde{x},\tilde{y})$ and $\varphi(\tilde{x},\tilde{y})$. If the invariance of the system under spin rotations is destroyed, i.e., the energy depends explicitly on φ ($L=$const but $S\neq$const), then only rotation solitons (8) with $\omega_p=0$ are possible. For a magnet with $L\neq$const but $S=$const (for example, $W_a=W(\theta)$ but $A_{xx}\neq A_{yy}$) the energy depends explicitly on χ. In this case only precession solitons (6) with $\omega_r=0$ are admissable.

If the model is such that only simultaneous rotation in spin and coordinate space is possible, then only the total angular momentum $J=S-L$ is conserved. Solitons in a physically nontrivial model of a magnet, where $S\neq$const and $L\neq$ const but $J=S-L=$const, were studied in Ref. 18 by direct numerical simulation. In polar coordinates the energy of such a magnet contains terms that depend on $\varphi-\chi$. It is obvious that in this case it is natural to use the substitution (10) with $\omega_p=\omega_r$, which yields a soliton of the type shown in Fig. 1c. For this case the quantity $\varphi-\chi$ is not explicitly time-dependent in a rotating coordinate system.

It is important to note that for the model (1) there always exists a simple centrosymmetric solution $\theta=\theta_0(r)$ and $\varphi=\nu\chi+\varphi_0$, where $\nu=\pm1,\pm2,\ldots$ is the topological charge and $\varphi_0=$const. In this case the expressions (6) and (8) are actually identical, Eq. (10), which contains $\delta W/\delta\varphi$, becomes an identity, and the function $\theta_0(r)$ is determined by an ordinary differential equation (see Refs. 1–3)

$$\frac{d^2\theta_0}{dx^2}+\frac{1}{x}\frac{d\theta_0}{dx}-\left(1+\frac{\nu^2}{x^2}\right)\sin\theta_0\cos\theta_0=(\Omega_p+\Omega_r)\sin\theta_0, \tag{11}$$

where $x=r/\Delta$, $\Omega_p=\omega_p/\omega_0$, $\Omega_r=\omega_r/\omega_0$, $\Delta=\sqrt{A/K}$ and $\omega_0=2\gamma K/M_0$ are, respectively, the characteristic length and frequency of the natural ferromagnetic resonance in an anisotropy field $H_a=2K/M_0$, and M_0 is the saturation magnetization. For a centrosymmetric soliton the difference between precession and rotation solitons vanishes. For this soliton with $\nu=1$ we have $S=L$ and $J=0$,[1–3] and only the sum of the frequencies ω_p and ω_r enters in Eq. (11).

If the anisotropy (last term on the left-hand side) is suppressed in this equation and $\Omega_{p,r}=0$, then the self-duality equation $d\theta_0/dx=-(\nu/x)\sin\theta_0$ is easily obtained. The solution of the latter equation gives a Belavin–Polyakov soliton (2) (in what follows we shall consider only the soliton with $\nu=1$, which has the lowest energy). Taking account of anisotropy effects, Eq. (11) can be easily integrated by the shooting method.

In summary, the structure of a centrosymmetric soliton can be easily investigated; see Refs. 1 and 2. However, we shall show that solitons without central symmetry are of greatest interest. Only such solitons exist for magnets with anisotropy in the basal plane. Moreover, such nonsymmetric solitons are realized even for the most symmetric model (1).

It is much more difficult to investigate soliton solutions which do not possess central symmetry. The structure of a soliton is determined by the two nonlinear partial differential equations (5), taking account of Eqs. (9), for the functions $\theta(r,\chi)$ and $\psi(r,\chi)$. There is no general method for analyzing the localized solutions of such equations and their stability. Soliton solutions can be constructed numerically by molecular-dynamics methods, but this requires a great deal of computer time.[18] Direct variational methods have turned out to be very effective for analyzing soliton structure and stability.[14]

3. METHOD OF ANALYSIS AND SOLITON STABILITY CONDITION

The equations for $\theta(r,\chi)$ and $\varphi(r,\chi)$ can be obtained by requiring that the auxiliary functional $\mathscr{L}\{\theta,\varphi\}=W-\omega_r L-\omega_p S$ be an extremum, where W is the energy of the ferromagnet. We start from the expression

$$\mathscr{L}\{\theta,\varphi\}=\int d^2x\left\{\frac{1}{2}A[(\nabla\theta)^2+\sin^2\theta(\nabla\varphi)^2]+\frac{1}{2}K\sin^2\theta+\Delta W(\theta,\varphi)-\frac{M_0}{\gamma}(1-\cos\theta)\times[\omega_p+\omega_r(\partial\varphi/\partial\chi)]\right\}. \tag{12}$$

Here the first two terms are the energy (1) in angular variables and the last term specifies the dynamic part. The term $\Delta W(\theta,\varphi)$ is a correction to the energy of a ferromagnet that lowers the dynamical symmetry of the magnet and describes the breakdown of a certain integral of motion.

We note that when the expressions (6), (9), or (10) are taken into account, the functional $\mathscr{L}$ becomes identical to the Lagrangian of a ferromagnet. Therefore the solutions of the form (6)–(9) to the Landau–Lifshitz equation are also extremals of the functional $\mathscr{L}$. On the other hand, it follows from Eq. (12) that a soliton corresponds to a conditional extremum of the energy for fixed L and S. The quantities ω_r and ω_p are Lagrange multipliers.

Therefore the analysis of the structure of a soliton reduces to searching for the extremals of the functional $\mathscr{L}\{\theta,\varphi\}$. For simplicity we shall assume that $\mathscr{L}$ depends not on the continuous functions $\theta(r,\chi)$ and $\varphi(r,\chi)$ but rather on a large but finite number of discrete parameters a_i, $i=1, 2,\ldots,n$. This simplication is not essential. Actually, we shall seek the extremum of $\mathscr{L}$ on a class of trial functions that depend on an arbitrary number of parameters a_1, a_2, $\ldots,a_n$, but for the general analysis performed in the present section the origin of the parameters a_i is not important. Specifically, such a situation arises when analyzing discrete models on a large but finite lattice. Moreover, for a system of finite size the a_i can be treated as coefficients in the expansion of the solutions in terms of a complete set of functions. Since far from a soliton the solutions decay exponentially, the parameters a_i with large i are unimportant.

If the state of the magnet is described by n parameters a_i, then $\mathscr{L}=\mathscr{L}(a_1,a_2,\ldots,a_n,\omega_p,\omega_r)$ and the condition of an extremum has the form of a system of n algebraic equations $\partial\mathscr{L}/\partial a_i=0$. Once the the solution $a_i^{(0)}$ is found, the integrals of motion E, S, and L characterizing a soliton with fixed ω_r and ω_p can be calculated. Eliminating the parameters ω_r and ω_p, it is possible to construct the function $E=E(L,S)$. It is easy to show that the relations

$$\frac{\partial E(L,S)}{\partial L}=\omega_r \quad \text{and} \quad \frac{\partial E(L,S)}{\partial S}=\omega_p \tag{13}$$

hold.

We note that the problem here is to search for an extremum, but not necessarily a minimum. Of course, a stable soliton should correspond to a minimum of the energy for fixed values of the integrals of motion L and S. However, in our method of Lagrange multipliers the auxiliary functional (Lagrangian) $\mathscr{L}$ may not have a minimum. Indeed, we shall verify below that dynamical solitons in ferromagnets (both stable and unstable) correspond to a saddle point of the functional $\mathscr{L}$. Thus, the question of the stability of solitons is very nontrivial. Fortunately, it can be analyzed in a general form, and the stability condition can be expressed in terms of the integral characteristics of solitons.

We employ Lyapunov's direct method (see Ref. 19) to analyze the stability. In this method a soliton is stable if there exists a Lyapunov functional $\Lambda\{\theta,\varphi\}$ such that 1) the functional is positive-definite near the soliton solution and 2) its time derivative, found taking account of the equations of motion, is negative or zero. We take the Lyapunov functional in the form of a combination of integrals of motion

$$\Lambda=\mathscr{L}+\frac{1}{2}B_1(L-L_0)^2 + B_2(S-S_0)(L-L_0)+\frac{1}{2}B_3(S-S_0)^2. \tag{14}$$

where L_0 and S_0 are the values of L and S for the soliton and B_1, B_2, and B_3 are constants. Then the condition 2) holds in the form $d\Lambda/dt=0$. The Lyapunov functional Λ in the form of a bilinear combination of integrals of motion has been chosen previously for analyzing the stability of nontopological magnetic solitons[19] and one-parameter rotation solitons in orthorhombic ferromagnets.[14] It was found that if $\Lambda<0$ holds somehwere, then the soliton is unstable. We have shown that this condition of instability also holds in the present case (the Chetaev functional describing instability is chosen in the same form as in Ref. 19). Therefore solitons are stable for $\Lambda>0$, and violation of this condition implies that they are unstable.

To analyze the stability condition (*i*) we shall investigate Λ for small deviations of the parameters a_i from $a_i^{(0)}$. Introducing $\alpha_i=a_i-a_i^{(0)}$ we write the value of the functional in the quadratic approximation in α_i as

$$\Lambda=\frac{1}{2}\sum_{i,k}\mathscr{L}_{ik}\alpha_i\alpha_k+\frac{1}{2}B_1\Big(\sum_i L_i\alpha_i\Big)^2 + B_2\Big(\sum_i L_i\alpha_i\Big)\Big(\sum_i S_i\alpha_i\Big)+\frac{1}{2}B_3\Big(\sum_i S_i\alpha_i\Big)^2, \tag{15}$$

where $\mathscr{L}_{ik}=\partial^2\mathscr{L}/\partial a_i\partial a_k$, $L_i=\partial L/\partial a_i$, and $S_i=\partial S/\partial a_i$. We reduce the matrix $\mathscr{L}_{ik}$ to the diagonal form $\mathscr{L}_{ik}=$ diag $(\lambda_1,\ldots,\lambda_n)$. The eigenvalues ε_i of the quadratic form Λ are given by the determinant of the system of linear equations for α_i:

$$(\varepsilon_i-\lambda_i)\alpha_i+\Big(\sum_j L_j\alpha_j\Big)(B_1L_i+B_2S_i) + (B_2L_i+B_3S_i)\Big(\sum_j S_j\alpha_j\Big)=0. \tag{16}$$

Premultiplying these equations by $L_i/(\lambda_i-\varepsilon_i)$ and $S_i/(\lambda_i-\varepsilon_i)$ and summing over i, we can write the equation for ε_i in the form of the condition for solvability of a system of two linear equations for $(\Sigma_jL_j\alpha_j)$ and $(\Sigma_jS_j\alpha_j)$. In what follows we shall need to consider large values $B_i\gg1$.[19,20] To lowest order in $1/B$ this condition can be represented in the form

$$F(\varepsilon)=\Big(\sum_i\frac{S_i^2}{(\varepsilon_i-\lambda)}\Big)\Big(\sum_i\frac{L_i^2}{(\varepsilon_i-\lambda)}\Big) -\Big(\sum_i\frac{L_iS_i}{(\varepsilon_i-\lambda)}\Big)^2=0. \tag{17}$$

The form of the function $F(\varepsilon)$ is virtually identical to that in the case of a single integral of motion,[14,19,20] specifically, $F(\varepsilon)\to+0$ as $\varepsilon\to-\infty$, and for values of ε close to λ_i (the

λ_i are eigenvalues of the matrix $\mathscr{S}_{ik}$) $F(\varepsilon)$ possesses poles of the form $F \simeq A/(\varepsilon-\lambda_i)$, $A>0$. Therefore the eigenvalues of $\Lambda(\varepsilon_i)$ lie between λ_i and λ_{i+1}. Hence if the matrix $\mathscr{S}_{ik}$ possesses two or more negative eigenvalues, one of the values satisfies $\varepsilon<0$ and the first condition of the Lyapunov theorem does not hold. In this case it is possible to construct the Chetaev functional[14] which means that the soliton is unstable. However, if only one eigenvalue λ_i is negative, then the stability is determined by the sign of $F(0)$: for $F(0)>0$ one value satisfies $\epsilon<0$, while for $F(0)<0$ all values satisfy ε are positive and the soliton is stable (see Fig. 1 of Ref. 14). We note that if all $\lambda_i \geqslant 0$, then all $\epsilon>0$ and the soliton is stable (but in our case it turned out that at least one eigenvalue satisfied $\lambda_i<0$; see below).

The quantity

$$F(0)=\left(\sum_i \frac{L_i^2}{\lambda_i}\right)\left(\sum_i \frac{S_i^2}{\lambda_i}\right)-\left(\sum_i \frac{L_i S_i}{\lambda_i}\right)^2$$

can be written in terms of the derivatives $\partial L/\partial\omega_r$ and $\partial S/\partial\omega_p$. Indeed, differentiating the relation $\partial\mathscr{L}/\partial a_i=0$ with respect to ω_p and ω_r one easily finds

$$\frac{\partial F}{\partial\omega_\alpha}=\sum\left(\frac{\partial F}{\partial a_i}\right)\left(\frac{\partial a_i}{\partial\omega_\alpha}\right),$$

where $F=S$ or L and $\alpha=r$, p. Then the sums in $F(0)$ can be easily expressed in terms of $\partial L/\partial\omega_\alpha$ and $\partial S/\partial\omega_\alpha$, and the stability condition becomes

$$\frac{\partial(L,S)}{\partial(\omega_p,\omega_r)}=\frac{\partial L}{\partial\omega_r}\frac{\partial S}{\partial\omega_p}-\frac{\partial S}{\partial\omega_r}\frac{\partial L}{\partial\omega_p}<0. \tag{18}$$

We note that general relations of the same structure as Eq. (18) arise in the problem of the stability of a moving soliton in a uniaxial antiferromagnet[20] (the integrals of motion are S and $\mathbf{P}$, $\partial L/\partial\omega_r \to \partial P_i/\partial v_i$, and $\mathbf{v}$ is the velocity of the soliton) as well as for two-parameter optical solitons.[21] Apparently, the relation (18) is of the same general character as the well-known condition of stability $\partial S/\partial\omega<0$ for one-parameter solitons; see Refs. 12, 14, and 19. We note that after simple algebraic transformations the condition $\partial S/\partial\omega<0$ is obtained from Eq. (18) in the case of one-parame te r centrosymmetric solutions, where $S\to L$.

Therefore time-periodic (or almost periodic for the model (1) of highest symmetry) soliton solutions are stable provided that one eigenvalue satisfies $\lambda_i<0$ and the condition (18) holds (as well as in the case where $\lambda_i>0$, but this case does not occur for the solitons considered in the present paper). Stability outside the class of periodic solutions requires a special analysis, which falls outside the scope of the present paper, especially since in the works known to us on dynamic solitons the authors confine their attention to the case of periodic solutions; see Refs. 1–3, 12, and 21.

4. CHOICE OF A TRIAL FUNCTION

The choice of a trial function is the decisive factor in any variational calculation. For the soliton problem a good decision would be to replace the energy (1) by its discrete analog, considering N classical spins occupying sites of a quite large lattice, and to choose the angles θ_k and φ_k for the spin s_k, $k=1,\ldots,N$, as the trial parameters. This approach for moving nontopological solitons was implemented in Ref. 17 with $N=249\times249$. However, calculations with a large number of parameters $n=2N$ require quite powerful computers, and in addition there arises an uncontrollable effect due to the discreteness. Specifically, the conservation of L certainly breaks down for discrete models (see below).

In the present model additional complexities arise because the desired extremum of $\mathscr{L}$ is not a minimum. Ordinarily, the standard minization programs seek an absolute minimum, so that they are inapplicable in the present case. Therefore to search for an extremum it is necessary to solve (as a rule, by iteration methods) a system of $n=2N$ nonlinear transcendental equations $\partial\mathscr{L}/\partial a_i=0$.

On the other hand, as we showed above, the condition of stability of a topological soliton can be written in terms of its global characteristics—the character of the dependence of the integrals of motion on the parameters ω_r and ω_p. Thus, there is hope that even quite simple trial functions will give good results provided that the trial function chosen admits a sufficiently wide class of disturbances that can be ''dangerous'' from the standpoint of soliton stability. These are, first and foremost, the change in the dimensions of a soliton and the elliptic deformation of its shape and the distribution of the magnetization vector in the xy plane for fixed $\theta(r,\chi)$.

For specific calculations we chose a trial function of the form

$$\tan\frac{\theta}{2}=\frac{R}{r}\exp\left(-\frac{r}{b}\right)(1+C_1\cos 2\chi),$$

$$\varphi=\chi+C_2\sin 2\chi+\varphi_0, \tag{19}$$

which depends on five trial parameters R, b, C_1, C_2, and φ_0 and gives a good approximation of the structure of a soliton.[22] Indeed, the function $\theta(r)$ (19) gives a description of the Belavin–Polyakov limit, adequate for $R\ll\Delta$, and also exponential decay of $\theta(r)$ for $r>\Delta$, where $\Delta=\sqrt{A/K}$ is a characteristic length (see Refs. 1–3). Therefore the function (19) describes well the form of the function $\theta(r)$; see below. The angular dependences agree with those obtained analytically in the limiting cases $R\gg\Delta$ and $R\ll\Delta$ or small asymmetry of the soliton, $C_{1,2}\ll1$.[22] The parameter C_1 controls the anisotropy of the function $\theta(r,\chi)$, i.e., elliptic distortions of soliton shape, and C_2 controls the anisotropy of the angular dependence $\varphi(\chi)$.

As noted above, the soliton stability criterion (18) is written in terms of the integral characteristics of the soliton. Therefore there is hope that analysis of this criterion using more accurate functions than Eq. (19) would produce only a small (to the extent that the deviations of $E(L)$ or $E(S)$ from their true values are small) displacement of the points of instability found using Eq. (19) and will not change the picture of the bifurcations of the solutions.

The system of equations $\partial\mathscr{L}/\partial a_i=0$ was solved by Newton's iteration method. The initial values of the parameters a_i were set manually. The two-dimensional integrals were calculated by a subroutine based on the recursive algorithm of the QUADREC subroutine.[23] Using the values

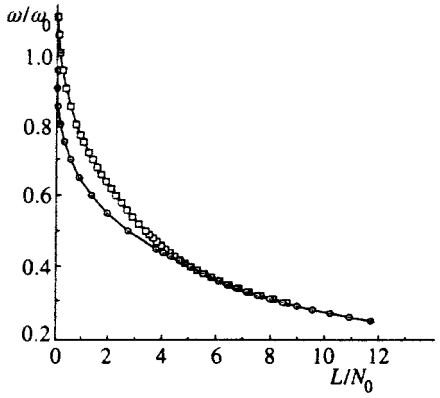

FIG. 2. $\omega(L)$ for a rotation soliton. The squares correspond to $\varepsilon=0.5$ and the circles to $\varepsilon=10^{-5}$.

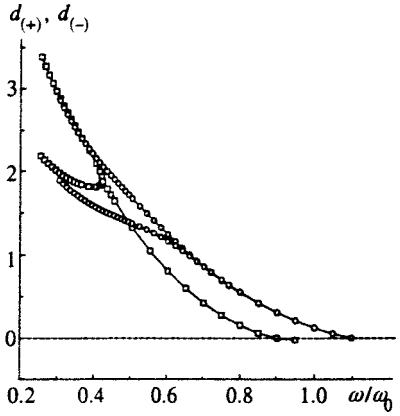

FIG. 3. $d_{(+)}$ and $d_{(-)}$ versus frequency for a rotation soliton in a ferromagnet with $\varepsilon=0.5$ (circles) and $\varepsilon=10^{-5}$ (squares).

found for $a_i^{(0)}$ we calculated the values of E, L, and S and constructed the functions $E(L,S)$ or $E(S)$, $E(L)$, and $E(J)$ (see below). The eigenvalues λ_i of the matrix $\mathscr{L}_{ik}=\partial^2\mathscr{L}/\partial a_i\partial a_k$ were calculated at the same time. It turned out that at least one eigenvalue λ_i is less than zero, i..e. a soliton always corresponds to a saddle point of $\mathscr{L}$. To analyze stability in the case of two-parameter solitons the quantity $\partial(L,S)/\partial(\omega_r,\omega_p)$ was also calculated.

5. ANALYSIS OF PERIODIC SOLITONS IN MODELS WITH ONE INTEGRAL OF MOTION

A specific investigation of solitons was performed both for the model (1) with the maximum possible symmetry and for less symmetric models which take account of additional terms that lower the symmetry of the problem [i.e., they destroy certain integrals of motion (4)]. It is convenient to begin the description of the results with the cases of less symmetric models where only the orbital angular momentum L or the total spin S or the total angular momentum $J=S-L$ is conserved. Two-parameter solitons for the model (1) of highest symmetry will be examined in the next section.

5.1. $L=$const, $S\neq$const. Let conservation of S break down because of the presence of magnetic anisotropy in the basal plane. For a specific analysis we choose an orthorhombic anisotropy of the form

$$\Delta W_1=\frac{1}{2}\varepsilon K m_y^2=\frac{1}{2}\varepsilon K\sin^2\theta\sin^2\varphi. \tag{20}$$

In this case L remains an integral of motion and rotation solitons of the form (8) are possible.

Taking account of the additional term (20), the lowest magnon frequency is $\omega_0\sqrt{1+\varepsilon}$. This value of the frequency is the maximum possible value in a soliton, i.e., rotation solitons exist only for $\omega_r<\omega_0\sqrt{1+\varepsilon}$. In the present subsection we shall drop the index r when describing rotation solitons.

A calculation showed that solitons exist for all values $0<\omega<\omega_0\sqrt{1+\varepsilon}$ (see Fig. 2). When the condition $\omega_0\sqrt{1+\varepsilon}-\omega_r\ll\omega_0$ is satisfied for any values of ε, the radius of the soliton is small, $R\ll\Delta$, and L is much less than the characteristic quantity N_0 ($N_0\gg1$ for $\Delta\gg a$),

$$N_0=2\pi s(\Delta/a)^2. \tag{21}$$

As $\omega\to\omega_0\sqrt{1+\varepsilon}$, this energy approaches the value $E_0=4\pi A$, equal to the energy of a Belavin–Polyakov soliton. The values of the asymmetry parameters C_1 and C_2 for $L\ll N_0$ remained small even for $\varepsilon\sim1$ (C_1 and $C_2<10^{-2}$ for $\varepsilon=0.5$ and $\omega>0.8\omega_0$, which corresponds to $L\leqslant0.03N_0$).

As the frequency decreases, the values of E and L increase, i.e., $d\omega/dL<0$; see Fig. 2. For a ferromagnet with $\varepsilon\neq0$ the parameters C_1 and C_2 likewise increase. For small ε these parameters depend nonlinearly on ε, i.e., for some value of ω/ω_0 (or L/N_0) the asymmetry of the soliton grows rapidly (see Fig. 3). Here and below we describe the asymmetry of the distribution of the magnetization **m** in a soliton using the two parameters $d_{(+)}$ and $d_{(-)}$, which correspond to the maximum and minimum sizes of the region where $\theta_0\geqslant\pi/2$. The relation of these parameters with the parameters of the trial function (19) for $\theta=\theta(r,\chi)$ is given by

$$d_\pm=R(1\pm C_1)\exp(-d_\pm/b).$$

For small ε ($\varepsilon\simeq10^{-4}-10^{-5}$) there arises an interesting effect that is helpful for understanding the properties of a soliton in the isotropic model (1). Specifically, for $\varepsilon\ll1$ and all $\omega\leqslant\omega_0$ almost centrosymmetric solitons with C_1, $C_2\ll1$ are easily found. These solutions satisfy $L(\omega)=|L_z|\simeq S(\omega)$. Then the function $E(S)$ is identical, to within quantities of order ε, to the function obtained previously by integrating the equation for θ_0.[1–3] This demonstrates the adequacy of the method and the trial function (19).

It turned out that the solution need not be single-valued. The program found a particular solution as a function of the choice of the initial values a_i. Specifically, for small ε (just as for $\varepsilon=0$) and $\omega\leqslant0.418\omega_0$, together with centrosymmetric solitons, asymmetric solutions with strongly different $d_{(+)}$ and $d_{(-)}$ were also found. We note that for a soliton with ε not small the difference $d_{(+)}-d_{(-)}$ depended continuously on ω, whereas for small ε the transition to an asymmetric soliton looks like a bifurcation (see Fig. 3).

It is important that for asymmetric solutions the value of $|L(\omega)|$ increased with decreasing ω much more rapidly than $S(\omega)$. Therefore the functions $E(S)$ and $S(L)$ differed fundamentally. If solitons with a fixed value of S are considered,

then the centrosymmetric solitons have the lowest energy. However, if L is fixed, then the asymmetric solitons have the lowest energy.

For all solitons considered, the derivative $dL/d\omega<0$, which is necessary for the solitons to be stable. The second condition—only one eigenvalue of the matrix $\mathscr{L}_{ik}$ can be negative—was satisfied only for a soliton whose energy is minimum for fixed L. Specifically, for $\varepsilon=0$ and a centrosymmetric soliton in a ferromagnet the second eigenvalue λ_2 is positive only for $\omega\geqslant 0.418\omega_0$. For $\omega\leqslant 0.418\omega_0$, when an asymmetric solution appears, the sign of λ_2 changes and the centrosymmetric soliton becomes unstable. Then for an asymmetric soliton only one eigenvalue is negative, and such a soliton is stable in its entire region of existence. Therefore the symmetry of rotation solitons can be lower than the symmetry of the model. This effect was discussed in the brief communication[14] in terms of spontaneous breaking of the symmetry of a centrosymmetric soliton under elliptic deformations and formation of a stable asymmetric soliton.

5.2. $L\neq$ const, $S=$ const. This case corresponds to a ferromagnet in which there is no magnetic anisotropy in the basal plane ($W_a=W_a(\theta)$ and does not depend on φ), but symmetry under spatial rotations is absent. Specifically, this situation always arises when switching from lattice models to continuum models. For example, for a square lattice, which is often used in numerical simulation, the nonconservation of L is due to invariants of the form $(\partial^2\mathbf{m}/\partial x\partial y)^2$. For an orthorhombic ferromagnet nonconservation of L could arise if the difference of the exchange interaction along the x and y axes is taken into account. Then terms of the form

$$(A_{xx}-A_{yy})(\partial\mathbf{m}/\partial x)^2 = (A_{xx}-A_{yy})\left\{\left[\left(\frac{\partial\theta}{\partial r}\right)^2+\sin^2\theta\left(\frac{\partial\varphi}{\partial r}\right)^2\right]\cos^2\chi + \frac{1}{r^2}\left[\left(\frac{\partial\theta}{\partial\chi}\right)^2+\sin^2\theta\left(\frac{\partial\varphi}{\partial\chi}\right)^2\right]\sin^2\chi\right\} \tag{22}$$

appear in the energy. The addition of a term with a simpler structure

$$\Delta J\sin^2\theta\sin^2\chi \tag{23}$$

to the energy of the magnet also gives the same effect. For the energy (1), taking account of the correction (22), the problem can be solved exactly by introducing the new variables $x'=x/\sqrt{A_{xx}}$ and $y'=y/\sqrt{A_{yy}}$. Then, in the polar coordinates introduced for the Cartesian coordinates x', y' ($r'=\sqrt{x'^2+y'^2}$, $\chi'=\tan^{-1}(y'/x')$) a solution of the problem is $\theta=\theta(r')$ and $\varphi=\nu\chi'+\omega t$. It is obvious that this solution is asymmetric in the initial physical coordinates x and y, the asymmetry being due strictly to the value of the parameter $(A_{xx}-A_{yy})/(A_{xx}+A_{yy})$. Therefore, in contrast to the case of a rotation soliton considered above, the soliton asymmetry parameter $d_{(+)}-d_{(-)}$ increases continuously with $(A_{xx}-A_{yy})/A_{xx}$ and spontaneous symmetry breaking does not arise. Our numerical analysis showed that the same situation also obtains for a model with a correction of the form (23).

For a square-lattice magnet the corresponding term destroying the conservation of L contains an additional small parameter $(a/\Delta)^2$. Since the effect of discreteness on soliton structure is small even for $\Delta\simeq 1.5a$,[24] in real weakly isotropic magnets with a square lattice the effects due to nonconservation of the orbital angular momentum with $S=$const should be weak.

5.3. $L\neq$const, $S\neq$const, $J=$const. This case must be viewed as the most physical in relation to continuum theories. However, its advantages for real models of two-dimensional magnets are not obvious. One of the physically interesting interactions possessing such symmetry is the long-range part of the magnetic dipole interaction, which does not reduce to renormalization of the magnetic anisotropy. Another example is the magnetoelastic interaction, which is described by terms of the form $\lambda m_i m_k(\partial u_i/\partial x_k)$, where $\partial u_i/\partial x_k$ is the distortion tensor.

Dynamic solitons with topological charge $\nu=2$, taking account of an additional term of the form $\delta(\nabla\cdot\mathbf{m})^2$, modeling the long-range part of the magnetic-dipole interaction, to the interaction energy (1), have been studied numerically in Ref. 18.

To analyze the effects of such a breaking of symmetry we chose a correction to the energy (1) of the form

$$\delta\sin^2\theta\sin^2(\varphi-\chi). \tag{24}$$

A term proportional to $\sin^2(\varphi-\chi)$ arises when $(\nabla\cdot\mathbf{m})^2$ is written in polar coordinates. This choice admits a direct comparison of effects due to an interaction of the form (24) with the previously considered interactions of the form (20) and (23). We shall discuss the results of an analysis of these models in greater detail.

First of all, it is obvious that solitons in the model (1), taking account of the correction (24), correspond to solutions of the form (9) with $\omega_r=|\omega_p|=\omega$. The superposition of two types of dynamics (precessional and rotational) with $\omega_r=|\omega_p|$ can be described as a rotation of a ''rigid'' soliton (see Fig. 1c) and as an oscillation of its form. Motion of this type (rotation of a ''rigid'' soliton pair) has been observed in numerical experiments.[18]

It is clear from analysis of Fig. 1c that in contrast to rotation solitons ($\omega_r\neq 0$, $\omega_p=0$) or precession solitons ($\omega_r=0$, $\omega_p\neq 0$), global dynamics (of the magnetization precession type) is absent in this case at a given point of the magnet. In other words, at a given point the magnetization undergoes only small oscillations around a definite average direction without a complete turn even for $\theta\simeq\pi/2$. For this reason, such solitons can be appropriately called oscillation solitons.

The difference of the properties of oscillation solitons and the rotational and precession solitons considered above is most clearly shown by analyzing the limiting case of a centrosymmetric soliton. For an oscillation soliton with $\omega_p=|\omega_r|$, there is no dynamics at all in this case and the solution degenerates into a static solution, while the centrosymmetric rotation and precession solitons are dynamic. It is obvious that solitons in the model $J\equiv S_z+L_z=S-L=$const and L, $S\neq$const cannot be centrosymmetric and are always characterized by a finite shape asymmetry.

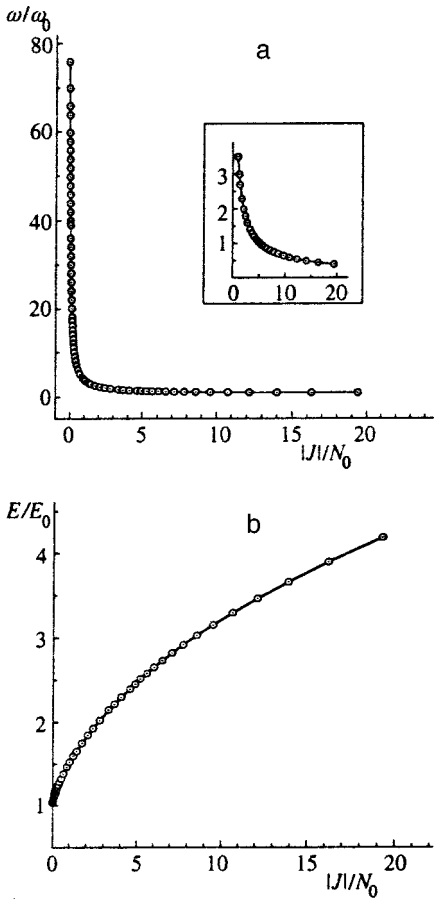

FIG. 4. $\omega(J)$ (a) and $E(J)$ (b) for an oscillation soliton with $\delta=0.5$.

Numerical analysis showed that in the present model there arises a new property that is absent for precessional and rotation solitons, specifically, the frequency of the soliton $\omega=|\omega_p|=\omega_r$ can be much higher than the lowest magnon frequency ω_g ($\omega_g \geqslant \omega_0$); see Fig. 4a. Frequencies $\omega \gg \omega_g$ correspond to very small solitons, $d_{(+)}$, $d_{(-)} \ll \Delta$, and energy close to E_0. The dependence of their energy on $|J|$ (see Fig. 4b) is similar to the corresponding dependence $E(L)$ for rotation solitons.[14]

As noted above, small-radius solitons in the models (1) and (24) correspond to very high frequencies $\omega \gg \omega_0$, which lie in the continous magnonspectrum. Here $d_{(+)}$ and $d_{(-)}$ do not differ very strongly. However, in contrast to rotation solitons the ratio $d_{(+)}/d_{(-)}$ does not decrease below 1.175, even for $\omega \simeq 75\omega_0$; see Fig. 5.

A soliton with $\omega \gg \omega_g$ and $d_{(+)} \sim d_{(-)}$ corresponds to small shape oscillations with small amplitude but quite high frequency. It is obvious that the existence of a soliton with $\omega > \omega_g$ is due to the fact, noted above, that the magnetization dynamics in an oscillation soliton is not global. In principle, such oscillations should give rise to magnon radiation in the continuous spectrum, which is responsible for relaxation of the soliton. However, direct computer simulation of the motion of vortex pairs in an easy-axis ferromagnet[6] with a gapless magnon dispersion law ($\omega_g=0$) has demonstrated that relaxation of solitons due to magnon emission is slow. Such relaxation effects should be small in our model also, where

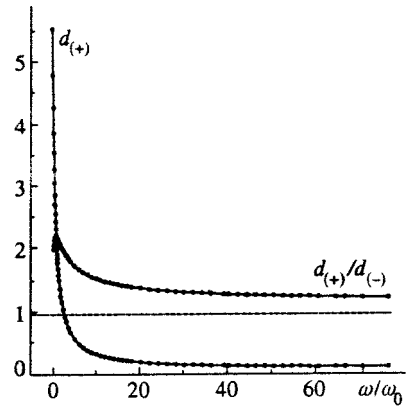

FIG. 5. Maximum size of the region of an oscillation soliton $d_{(+)}$, where $\theta \geqslant \pi/2$, and the ratio $d_{(+)}/d_{(-)}$ versus the frequency of magnetization oscillations in the soliton.

magnons have a finite gap ω_g but $\omega > \omega_g$. Such relaxation effects were also negligibly small in the study of soliton dynamics in an easy-plane ferromagnet with an amplitude that is not small.[18]

As the soliton frequency decreases to values below ω_0, the energy of thesoliton grows to values of the order of several E_0. The difference of the quantities $d_{(+)}$ and $d_{(-)}$ also increases (see Fig. 5), i.e., for $\omega \leqslant \omega_0$ a soliton in the model (1) and (24) is an excited state with a quite high energy and sharply asymmetric magnetization distribution, far from centrosymmetric. For $\omega < 0.3\omega_0$ the value of the parameter C_2 rises above $\pi/2$ and, by virtue of Eq. (19), the function $\varphi(x)$ becomes nonmonotonic. In this case the second value of λ changes sign and the soliton becomes unstable.

In summary, solitons in ferromagnetic models with $L=$ const or $J=$const demonstrate two different types of behavior. Rotation solitons ($L=$const) can be almost centrosymmetric and they can also be characterized by shape asymmetry that is not small. They always correspond to slow magnetization dynamics (the frequency ω_r lies below the gap in the magnon spectrum).

Oscillation solitons in a ferromagnet with $L \neq$const and $S \neq$const but $J=$const with the same values of the energy are more asymmetric than rotation solitons. They can be characterized by quite rapid variations of the magnetization $\omega \gg \omega_0$, which in this case can be clearly represented as small shape oscillations of the soliton. Such solitons can also exist as low-frequency solitons ($\omega \leqslant \omega_0$), in which case they are very asymmetric and have a quite high energy.

The differences of the ''oscillational'' ($|\omega_p|=\omega_r$) dynamics from precessional or rotational dynamics appear for two-parameter solitons in the model (1) with the maximum possible symmetry, for which $L=$const and $S=$const.

6. TWO-PARAMETER SOLITONS IN A MODEL WITH TWO INTEGRALS OF MOTION

As noted above, for the model (1) solitons with a complicated, generally speaking, almost periodic magnetization dynamics are possible. For them, the variation of the magnetization in a stationary coordinate system is characterized by two independent frequencies ω_r and ω_p.

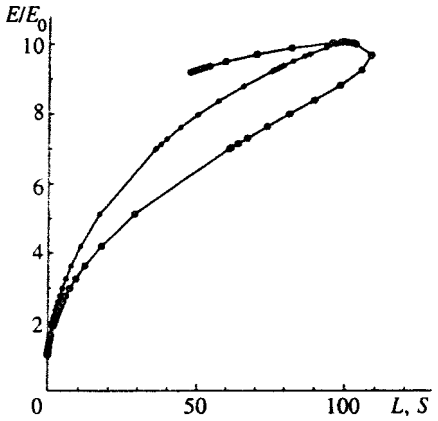

FIG. 6. Energy of a soliton versus the values of the integrals of motion L,S (in units of N_0) for a soliton with $\Delta\omega=0.1\omega_0$. The filled circles represent $E(S)$, and the open circles represent two branches of the function $E(L)$.

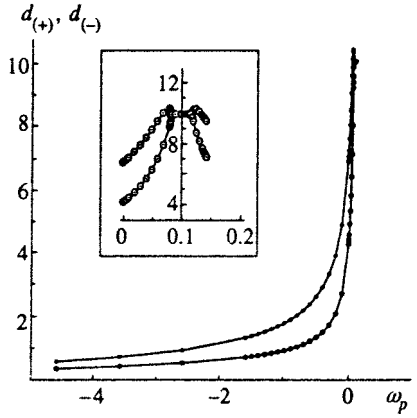

FIG. 7. Anisotropy parameters $d_{(+)}$ and $d_{(-)}$ of a soliton for $\Delta\omega=0.1\omega_0$.

The model (1) admits centrosymmetric solitons for which only the sum of frequencies $\Delta\omega=\omega_p+\omega_r$ is meaningful; see Eq. (11). In the general case, however, a soliton is not centrosymmetric. We shall consider the properties of solitons with the frequencies ω_r and ω_p varying so that $\Delta\omega=\omega_p+\omega_r=$ const. Analysis of such one-parameter families of solutions clearly demonstrates that the properties of solitons solitons change when the frequencies ω_r and ω_p change.

We note first that soliton solutions exist only for $\Delta\omega<\omega_0$. This is understandable, since $\Delta\omega$ determines the global dynamics of the magnetization in a soliton, i.e., $\Delta\omega$ plays the same role as ω_p for a precession soliton or ω_r for a rotation soliton. For small values $\Delta\omega\ll\omega_0$ there exist soliton solutions with large sizes ($d_{(+)}$, $d_{(-)}\gg\Delta$), and for $\Delta\omega$ close to ω_0 only centrosymmetric solitons are observed.

Just as for oscillation solitons, the frequencies ω_r and $|\omega_p|$ can be quite high (we observed solutions with $\bar{\omega}=(\omega_r-\omega_p)/2$ up to $100\,\omega_0$). For large values of $\bar{\omega}$ (but $\Delta\omega<\omega_0$) the properties of a soliton are virtually independent of $\Delta\omega$ and the same as for an oscillation soliton with $\omega_r=|\omega_p|\gg\omega_0$: the energy of the soliton is close to E_0, the size of the soliton is small, and $|L_z|=L\simeq S\ll N_0$. In this range the values of $d_{(+)}$ and $d_{(-)}$ are close. This is understandable from the analysis in the preceding section: for $\bar{\omega}\gg\Delta\omega$ the quantity $\bar{\omega}$ plays the same role as the frequency ω for oscillation solitons, and it describes small rapid oscillations of the shape of the soliton. But for two-parameter solitons these oscillations are superposed on a slow spin precession with the frequency $\Delta\omega$.

Decreasing the frequencies ω_r and $|\omega_p|$ so that $\Delta\omega$ remains constant, we observe an increase in the energy of the soliton and the values of L and S; see Fig. 6. At the same time, the soliton becomes more asymmetric (for $\Delta\omega=0.1\omega_0$ the maximum value satisfies $d_{(+)}/d_{(-)}\simeq 2.05$ for $\omega_r=0.7\omega_0$, $\omega_p=-0.6\omega_0$). However the ratio $d_{(+)}/d_{(-)}$ grows more slowly than the size of the soliton; see Fig. 7. In the region $|\omega_p|$, $\omega_r<\omega_0$ the characteristic values of $d_{(+)}$ and $d_{(-)}$ vary by one order of magnitude.

When the value of $\bar{\omega}$ becomes equal to $\Delta\omega$, the asymmetry of the soliton starts to decrease rapidly, and its average size increases. This fact has a quite obvious explanation. In this region the term related to S in the dynamical part of the Lagrangian $\mathscr{L}$ starts to dominate, and the soliton behaves as a precession soliton in the model with $S=$ const and $L\neq$ const. For $\omega_r\ll|\omega_p|\simeq\Delta\omega$ the central symmetry of the soliton is restored abruptly (see Fig. 7). For $\omega_p\gtrsim\Delta\omega$ only centrosymmetric solitons are observed. It is obvious that the magnetization dynamics in them does not depend on ω_p and ω_r, and their parameters and structure depend only on the value of $\Delta\omega$.

Over the entire range of variation of the parameters ω_p and ω_r the solitons described above (both asymmetric and centrosymmetric, arising for small ω_r) are stable in accordance with the criteria obtained above. Specifically, for all of them only one eigenvalue satisfies $\lambda_i<0$ and $\partial(L,S)/\partial(\omega_r,\omega_p)<0$. Besides these soliton states, centrosymmetric solitons are also present in the region of existence of the asymmetric solitons. However they are unstable (for them two eigenvalues of the matrix $\mathscr{L}_{ik}$ are negative). Therefore, here the situation is the same as for rotation solitons with $\varepsilon\ll 1$: if an asymmetric soliton exists, then it is stable. For the model (1) this can be described as spontaneous breaking of the symmetry of the soliton.

As ω_p increases further, asymmetric solitons reappear. This soliton branch corresponds to large values of the parameters C_1, $C_2\geqslant 1$. They have λ_1, $\lambda_2<0$, and they are unstable. It is interesting to note that for these unstable solitons the function $E(S)$ is the same as for stable solitons, so that these two branches coincide in the plot of $E=E(S)$; see Fig. 6. The main qualitative difference between these two branches, which are separated by a region where stable centrosymmetric solitons exist, consists in the ratio of L and S. Stable solitons have $|L_z|>S$, while for unstable solitons $S>|L_z|$ (see Fig. 8). We note that for stable rotation solitons it was always found that $|L_z|$ is greater than the average value of S). The difference between the branches of stable and unstable asymmetric solitons is not seen in the function $d_{\pm}(\omega)$ (see inset in Fig. 7), but it is clearly manifested in the function $E(L)$. Two branches of this curve with a characteristic bifurcation point at $E=10.003E_0$ and $|L|=S=99.815N_0$ are clearly seen in the function $E(L)$ (see Fig.

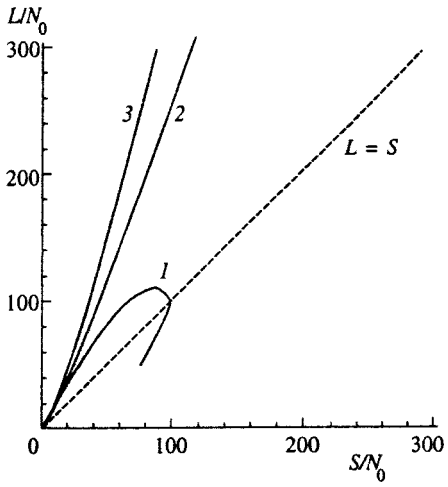

FIG. 8. L versus S for solitons with different values of $\Delta=\Delta\omega/\omega_0$. The curve *1* corresponds to $\Delta=0.1$, *2* — $\Delta=0.01$, and *3* — $\Delta=-0.1$.

6). This point corresponds to a centrosymmetric soliton whose structure depends only on $\omega_p+\omega_r$. At this point the energy E of the soliton and the z projection of the spin S are maximum for a given value of $\Delta\omega$.

The soliton properties described above were discussed for a particular one-parameter family of solitons with $\Delta\omega=0.1\omega_0$. Our analysis showed, however, that the qualitative behavior remains the same for other not very large positive values of $\Delta\omega$. Specifically, for $\Delta\omega\leqslant 0.5\omega_0$ a transition occurs from asymmetric to centrosymmetric solitons as ω_p decreases to $\Delta\omega$. For all $\Delta\omega>0$ there exists a maximum possible $E(\Delta\omega)$ that corresponds to a centrosymmetric soliton. Obviously, this value equals the energy of a centrosymmetric precession soliton with frequency $\omega=\Delta\omega$. For such $\Delta\omega$ there also exist frequencies for which $|L|<S$ and the corresponding solitons are unstable.

The situation changes for sufficiently large $\Delta\omega$ ($\Delta\omega\simeq\omega_0$) and negative $\Delta\omega$. For $\Delta\omega$ close to ω_0 we were able to find only centrosymmetric solutions. Since the corresponding solutions are stable, it can be concluded that asymmetric solitons are absent in this case. Actually, in this range of $\Delta\omega$ the solutions are one -parameter solutions and to each value of $\Delta\omega$ there corresponds only one value of S, $|L_z|=S$ and energy E. In the (L,S) plane these solitons are described by points on the straight line $L=S$ that lie in the region with small L, $S<N_0$.

Solutions with $\Delta\omega=\omega_r+\omega_p<0$ are never centrosymmetric, since centrosymmetric precession solitons in the model (1) exist only for $0<\omega<\omega_0$; see Refs. 1–3. For $\Delta\omega<0$ the shape asymmetry of the soliton is always larger than for solitons with the same energy E and $\Delta\omega>0$. In the (L,S) plane solitons with fixed $\Delta\omega<0$ correspond to an open curve lying above the line $L=S$ (see Fig. 8) and nowhere intersecting it.

7. CONCLUSIONS

We shall discuss the general mechanisms responsible for two-dimensional topological solitons, the properties of the solitons, and the possible manifestations of these solitons in the properties of various quasi-two-dimensional ordered systems.

Stable stationary topological $2D$ solitons exist for a wide class of models. The necessary condition is the existence of at least one angular momentum integral of motion (L, S, or J). In models with one integral of motion a soliton corresponds to periodic magnetization dynamics.

In the most highly symmetric model of a ferromagnet where both L and S are conserved two-parameter solitons are possible, the parameters being the precession frequency ω_p and the rotation frequency ω_r of the soliton. In these solitons the magnetization dynamics in a stationary coordinate system such that $q\omega_p\neq p\omega_r$ (p and q are integers) is described by an almost periodic function of time.

Solitons in models with one integral of motion can manifest quite diverse properties. The shape asymmetry of precession solitons in models with $S=$const is dictated by the intensity of the interaction that destroys the conservation of L. In an almost isotropic model such solitons are close to centrosymmetric. For them, just as for rotation solitons in models with $L=$const and $S\neq$const only a low-frequency magnetization dynamics is characteristic, ω_p, $\omega_r<\omega_g$, where ω_g is the lowest magnon frequency in the linear theory.

The shape asymmetry for rotation solitons is not related to the intensity of the interactions that destroy the symmetry of the model. However, for a very small magnetic anisotropy in the basal plane close to centrosymmetric solitons do exist and they are stable for energies $E>2.3E_0$. Even for a very small anisotropy in the basal plane, solitons with shape asymmetry that is not small are stable.

Oscillation solitons in models where the z projection of the total angular momentum J is conserved (provided that $L\neq$const and $S\neq$const) differ from precessional and rotation solitons primarily through having the magnetization dynamics that can be high-frequency ($\omega>\omega_0$). The shape asymmetry of these solitons is large for the low-frequency case ($\omega<\omega_0$), but it is not very small even at high frequencies, where $\omega\gg\omega_0$ and $J\ll N_0$.

The same high-frequency dynamics also appears for two-parameter solitons in the most highly symmetric model with $L=$const and $S=$const. We note that this circumstance could be important for analyzing the contribution of solitons regarded as nonlinear thermal excitations to the thermodynamics and the response function of quasi-two-dimensional uniaxial magnets. At low temperatures ($T\leqslant J\sim A$) mostly small solitons with $d_{(+)}\simeq d_{(-)}\ll\Delta$ and $E\simeq E_0$ are excited. For this reason, solitons with rapid oscillations should exist in a soliton gas, and these solitons (in contrast to those studied previously in Refs. 8 and 9) should make a specific contribution to the response function of the ferromagnet through high-frequency peaks. A detailed analysis of this problem falls outside the scope of the present paper.

Analysis of the most highly symmetric model with a purely isotropic basal plane (with respect to both the spatial and spin rotations) showed that the two-parameter solitons in this model can be asymmetric. As a rule, solitons with low energy, $E\simeq E_0$, are characterized by the presence of rapid

oscillations. Therefore, in this physically interesting model solitons with $R \ll \Delta$ and $E \simeq E_0$ can possess an asymmetric shape and can be characterized by rapid magnetization oscillations. Solitons with a higher energy (E greater than several E_0) are characterized by slow magnetization dynamics, ω_p, $\omega_r \simeq \omega_0$. They can be either asymmetric or centrosymmetric, depending on the values of ω_r, ω_p.

Two integrals of motion, L_z and S, exist in the model (1). When two independent integrals of motion L and S exist, the question arises as to which values of these integrals are admissable for a soliton, i.e., for what values of L and S does a stable soliton exist? The answer to this question can be obtained by analyzing the curves corresponding to $\Delta\omega =$ const in the (L,S) plane; see Fig. 8.

In this plane the straight line $L=S$ corresponds to centrosymmetric solitons. It corresponds to all $\Delta\omega$ close to ω_0 (in the limit $\Delta\omega \to \omega_0$ L and S are small and $E \simeq E_0$) as well as the points corresponding to stable centrosymmetric solitons for small $0 < \Delta\omega \ll \omega_0$ ($L,S \gg N_0$, $E \gg E_0$).

In the region $0 < L < S$ solitons are unstable, and we do not discuss them; see the discussion at the end of the preceding section. The regions where $L > S$ correspond to stable solitons, obtained for both small $\Delta\omega > 0$ and negative $\Delta\omega$.

These curves either intersect the line $L=S$, crossing into the region $L<S$ and forming a characteristic lobe (for $\Delta\omega > 0$), or they recede to infinity, remaining above the straight line $L=S$. A characteristic feature of all these curves $\Delta\omega =$ const is that for $L,S \geqslant N_0$ and small values of $\Delta\omega$ the curves lie much higher, i.e., the ratio L/S with fixed S increases with decreasing $\Delta\omega$. Therefore it can be asserted that for such values of L and S the half-plane $L>S$ is covered by the corresponding curves and a soliton with a fixed value $L > S \geqslant N_0$ can be constructed with an appropriate choice of $\Delta\omega$ and one of the frequencies ω_p and ω_r. The region of small $L,S \ll N_0$, where curves with different values of $\Delta\omega$ pass close to one another for both $\Delta\omega > 0$ and $\Delta\omega < 0$, is an exception. Hence it follows that solitons with $L \gg S$ but $L,S \ll N_0$ probably do not exist. This is understandable, considering that small values of L and S correspond to small soliton radii, $R \ll \Delta$; in this case the soliton shape anisotropy cannot be large, and $L \simeq S$.

This work was supported in part by the Ukrainian Foundation for Basic Research (Project 2.4/27).

*)E-mail: vbaryakhtar@bitp.kiev.ua

[1] A. M. Kosevich, B. A. Ivanov, and A. S. Kovalev, Phys. Rep. **194**, 117 (1990).
[2] V. G. Bar'yakhtar and B. A. Ivanov, Soviet Scientific Reviews, Sec. A — Phys., edited by I. M. Khalatnikov, **16**, No. 3, (1992).
[3] B. A. Ivanov and A. K. Kolezhuk, Fiz. Nizk. Temp. **21**, 355 (1995) [Low Temp. Phys. **21**, 275 (1995)].
[4] V. L. Berezinskiĭ, Zh. Éksp. Teor. Fiz. **59**, 907 (1971) [Sov. Phys. JETP **32**, 493 (1971)].
[5] J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973).
[6] A. R. Völkel, F. G. Mertens, A. R. Bishop, and G. M. Wysin, Phys. Rev. B **43**, 5992 (1991).
[7] F. Waldner, J. Magn. Magn. Mater. **31-34**, 1203 (1983); ibid. **54–57**, 873 (1986).
[8] C. E. Zaspel, T. E. Grigerate, and J. E. Drumhellar, Phys. Rev. Lett. **74**, 4539 (1995); C. E. Zaspel and J. E. Drumhellar, Int. J. Mod. Phys. **10**, 3649 (1996).
[9] B. A. Ivanov and A. K. Kolzhuk, Fiz. Nizk. Temp. **16**, 335 (1990) [Sov. J. Low Temp. Phys. **16**, 184 (1990)]; A. K. Kolezhuk and D. V. Filin, Fiz. Nizk. Temp. **20**, 1267 (1994) [Low Temp. Phys. **20**, 992 (1994)].
[10] A. A. Belavin and A. M. Polyakov, JETP Lett. **22**, 245 (1975); A. M. Perelomov, Usp. Fiz. Nauk **134**, 577 (1981) [Sov. Phys. Usp. **24**, 645 (1981)].
[11] S. L. Sondhi, A. Karlhede, S. A. Kivelson, and E. H. Rezayi, Phys. Rev. B **47**, 16419 (1993).
[12] V. G. Makhankov, Phys. Rep. **35**, 1 (1978).
[13] B. A. Ivanov, JETP Lett. **56**, 118 (1992).
[14] A. A. Zhmudskiĭ and B. A. Ivanov, JETP Lett. **65**, 945 (1997).
[15] N. Papanicolaou, in *Singularities in Fluids, Plasmas, and Optics*, edited by Gaflisch (NATO ASI Series, Springer, 1992).
[16] B. A. Ivanov, JETP Lett. **58**, 389 (1993).
[17] N. R. Cooper, Phys. Rev. Lett. **80**, 4554 (1998).
[18] N. Papanikolaou and W. J. Zakrzewski, Physica D **80**, 225 (1995).
[19] B. A. Ivanov and A. L. Sukstanskii, Solid State Commun. **50**, 523 (1984).
[20] I. V. Bar'yakhtar and B. A. Ivanov, Zh. Éksp. Teor. Fiz. **85**, 328 (1983) [Sov. Phys. JETP **58**, 190 (1983)].
[21] A. V. Buryak, Yu. S. Kivshar, and S. Trillo, Phys. Rev. Lett. **77**, 5210 (1996).
[22] A. A. Zhmudskiĭ and B. A. Ivanov, Fiz. Nizk. Temp. **22**, 446 (1996) [Low Temp. Phys. **22**, 347 (1996)].
[23] A. N. Berlizov and A. A. Zhmudsky, Preprint KINR-98-4 (Institute for Nuclear Research, Kiev, 1998).
[24] B. A. Ivanov, A. K. Kolezhuk, and G. M. Wysin, Phys. Rev. Lett. **76**, 511 (1996).

Translated by M. E. Alferieff