Contribution of the massive photon decay channel to neutrino cooling of neutron stars

D. N. Voskresensky*)

Moscow Institute for Physics and Engineering, 115409 Moscow, Russia;
Gesellschaft für Schwerionenforschung, D-64291 Darmstadt, Deutschland

E. E. Kolomeitsev†) and B. Kämpfer

Institut für Kern und Hadronenphysik, Forschungszentrum Rossendorf, D-01314 Dresden, Deutschland;
Institut für Theoretische Physik, TU Dresden D-01062 Dresden, Deutschland

(Submitted 15 August 1997)
Zh. Éksp. Teor. Fiz. **114**, 385–397 (August 1998)

We consider massive photon decay reactions via intermediate states of electron–electron-holes and proton–proton-holes into neutrino–antineutrino pairs in the course of neutron star cooling. These reactions may become operative in hot neutron stars in the region of proton pairing where the photon due to the Higgs–Meissner effect acquires an effective mass m_γ that is small compared to the corresponding plasma frequency. The contribution of these reactions to neutrino emissivity is calculated; it varies with the temperature and the photon mass as $T^{3/2}m_\gamma^{7/2}\exp(-m_\gamma/T)$ for $T<m_\gamma$. Estimates show that these processes appear as extra efficient cooling channels of neutron stars at temperatures $T\simeq 10^9-10^{10}$ K. © *1998 American Institute of Physics.* [S1063-7761(98)00108-5]

1. INTRODUCTION

The EINSTEIN, EXOSAT and ROSAT observatories measured surface temperatures of certain neutron stars and put upper limits on the surface temperatures of others (see Ref. 1 and further references therein). Data on the supernova remnants in 3C58, the Crab, and RCW103 indicate rather slow cooling, while the data for Vela, PSR 2334+61, PSR 0656+14, and Geminga point to significantly more rapid cooling. In the so-called standard scenario of neutron star cooling, the most important channel up to temperatures $T\leqslant 10^8-10^9$ K corresponds to the modified URCA process $nn\to npe\bar{\nu}$. Rough estimates of its emissivity were first made in Ref. 2. Friman and Maxwell[3] recalculated emissivity of this process in a model, in which the nucleon-nucleon interaction is treated with the help of slightly modified free one-pion exchange. Their result for emissivity, $\varepsilon_\nu^{\rm FM}$, proved to be an order of magnitude higher than previously obtained. The value $\varepsilon_\nu^{\rm FM}$ was used in various computer simulations resulting in the standard cooling scenario; see Ref. 4, for example. Subsequent works[5–7] took in-medium effects into account in *NN*-interaction, showing that emissivity of the modified URCA process depends heavily on neutron star mass. For stars of more than one solar mass, the resulting emissivities turned out to be substantially higher than the values given by $\varepsilon_\nu^{\rm FM}$.

These and other in-medium effects were recently incorporated in the computer code[8] leading to a new scenario of neutron star cooling. For low-mass stars numerical results of the new and standard scenarios more or less coincide. In the present work, we continue to look for enhanced reaction channels. To demonstrate the efficiency of new reaction channels, we compare the results with emissivity $\varepsilon_\nu^{\rm FM}$, which dominates cooling in the standard scenario over the temperature range under consideration.

Besides the modified URCA process, the standard scenario numerical codes also include neutron and proton bremsstrahlung processes $nn\to nn\nu\bar{\nu}$ and $np\to np\nu\bar{\nu}$, which in all models lead to a somewhat smaller contribution to emissivity than the modified URCA process.[3,5,6,9] Also included are processes that contribute to emissivity in the neutron star crust. These are plasmon decay $\gamma_{pl}\to\nu\bar{\nu}$,[10,11] electron bremsstrahlung on nuclei $eA\to eA\nu\bar{\nu}$,[11–13] electron-positron annihilation $ee^+\to\nu\bar{\nu}$,[14,15] and photon absorption by electrons $\gamma e\to e\nu\bar{\nu}$.[15–17] Numerical simulations show that the latter two processes contribute only negligibly to the crust neutrino emissivity at the temperatures under discussion in this paper and they always contribute negligibly to the full neutron star's emissivity; see Fig. 7 of Ref. 11.

When the temperature decreases, it is energetically favorable for neutrons to pair in the neutron star interior and inner crust and for the protons to pair in the star's interior. In a system with nucleon pairing, the emissivity of the modified URCA process is suppressed by a factor $\exp[-(\Delta_n+\Delta_p)/T]$,[3] where Δ_n and Δ_p are the respective neutron and proton gaps, defined by

$$\Delta_i(T)=\Delta_i(0)\,\frac{T_{c,i}-T}{T_{c,i}}\,\theta(T_{c,i}-T)$$

(here $\theta(x)$ is the Heaviside step function, $i=\{p,n\}$, and $T_{c,i}$ is the corresponding critical temperature for nucleon pairing). At temperatures $T\ll T_{c,p}$, $T_{c,n}$ the process becomes marginal. Nevertheless, this star's interior process still dominates those of crust cooling up to temperatures $T\sim 10^8-10^9$ K, depending on the values of the gaps; see

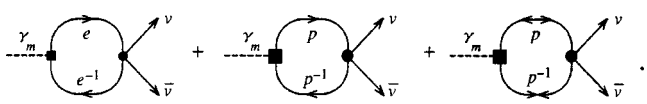
Fig. 7 of Ref. 11. For $T \leqslant (1-3)\cdot 10^8$ K cooling in the standard scenario is largely dominated by the photon emission from the neutron star surface.

In the present work we look for more efficient cooling processes at $T < T_{c,p}$, $T_{c,n}$. We analyze photon decay into neutrino-antineutrino pairs. The related processes $\gamma e \to e \nu \bar{\nu}$ and $\gamma p \to p \nu \bar{\nu}$ turn out to be suppressed by several orders of magnitude compared to those under discussion, due to the lack of free final states in degenerate fermionic systems, and are therefore not considered here. The contribution of photon decay via electron–electron-hole intermediate states for the case of a normal electron plasma in white dwarfs and neutron star crusts has been calculated by several authors (see Ref. 10 for further references). In an ultrarelativistic electron plasma, a photon acquires an effective in-medium plasmon dispersion law with a gap equal to the electron plasma frequency $\omega_{pl} \simeq 2e\mu_e/\sqrt{3\pi}$, where e is the electron charge and μ_e denotes the electron chemical potential (we employ units with $\hbar = c = 1$). Therefore, the contribution to emissivity of the cited process is suppressed by a factor $\exp(-\omega_{pl}/T)$. Nevertheless, in white dwarfs and neutron star crusts, the electron density is not too high, and the process is still effective. In neutron star interiors, the electron density ρ_e is equal to the proton density ρ_p by virtue electrical neutrality, and along with β stability one obtains a relation for the total density

$$\rho_e = \rho_p \simeq 0.016\,\rho_0 \left(\frac{\rho}{\rho_0}\right)^2, \tag{1}$$

where $\rho_0 \simeq 0.17$ fm^{-3} denotes the nuclear saturation density, and we use the values of the neutron and proton Fermi momenta,[3] $p_{Fn} \simeq 340(\rho/\rho_0)^{1/3}$ MeV and $p_{Fp} = \mu_e \simeq 85(\rho/\rho_0)^{2/3}$ MeV. Thus, at typical densities for neutron star interiors $\rho \gtrsim \rho_0$, the value of the electron plasma frequency is high, e.g., $\omega_{pl}(\rho_0) \approx 4.7$ MeV for $\rho \simeq \rho_0$, and at temperatures $T < T_{c,n}$, $T_{c,p} < \omega_{pl}$ the process $\gamma_{\omega_{pl}} \to e e^{-1} \to \nu \bar{\nu}$, where the superscript -1 denotes the hole, is strongly suppressed. We therefore seek another process that can contribute to rapid cooling.

We exploit the fact that, contrary to a normal electron plasma, in superconducting proton matter, due to the Higgs–Meissner effect, the photon acquires an effective mass that is small compared to the plasmon frequency. In the region of proton pairing at $T < T_{c,p}$, we therefore find that new decay processes of massive photons (γ_m) via electron–electron-hole (ee^{-1}) and proton-proton-hole (pp^{-1}) intermediate states to neutrino–antineutrino pairs, $\gamma_m \to ee^{-1} + pp^{-1} \to \nu_l \bar{\nu}_l$, $l = \{e, \mu, \tau\}$, can dominate neutron star cooling at certain temperatures. These processes are determined by the diagrams



In the first diagram, the solid lines in the loop are related to Green's functions of nonsuperfluid relativistic electrons. In the second and third diagrams, the solid lines in the loops correspond to superconducting nonrelativistic protons. The distinct orientations of arrows indicate that the second diagram is calculated with so-called ''normal'' Green's functions $\rightarrow\!\!-$, which become the usual Green's functions for normal Fermi liquids in the limit $\Delta_p \to 0$. In contrast, the third diagram is built up with the ''anomalous'' Green's functions $\leftrightarrow$ and $\rightarrow\!\!-\!\!\leftarrow$, which are proportional to the proton gap. Therefore the contribution of the third diagram vanishes for $\Delta_p \to 0$. The fat vertices in the second and third nucleon diagrams include nucleon–nucleon correlations.

The contribution to neutrino production matrix elements of the third diagram and terms proportional to the gap in the second diagram is as small as $(\Delta_p/\epsilon_{Fp})^2 \ll 1$ for $T < T_{c,p} \ll \epsilon_{Fp}$ (here ϵ_{Fp} is the proton Fermi energy), compared to the contribution of the second diagram calculated with the Green's functions of the normal Fermi liquid. To this same accuracy, we drop the third diagram and use the Green's functions of protons for the normal Fermi liquid[1)] in the second diagram. We thus calculate emissivity according to the first two diagrams, assuming $\Delta_p = 0$ in the second diagram but taking into account that the photon dispersion relation is changed due to proton superconductivity.

Our paper is organized as follows. In Sec. 2 we show that in the region of proton superconductivity due to the Higgs–Meissner effect, the photon spectrum is rearranged, and instead of the plasmon gap the photon acquires a mass, which is now determined by the density of paired protons. In Secs. 3 and 4 we demonstrate the efficiency of these new processes in the course of neutron star cooling. The emissivity corresponding to the above diagrams is calculated and compared with emissivity of the standard URCA process and photon emissivity from the neutron star surface. In Sec. 5 we detail our conclusions.

2. PHOTON SPECTRUM IN THE SUPERCONDUCTING PHASE

As is well known,[18] the photon spectrum in superconducting matter and in a normal plasma are substantially different. In the superconducting matter considered here, we deal with two subsystems. The normal subsystem contains electrons and nonpaired protons and neutrons, which are present to some extend at finite temperatures. The superfluid subsystem contains paired protons and neutrons. In the presence of a superconducting proton phase, normal currents associated with both electrons and residual nonpaired protons are fully compensated by the corresponding response of the superconducting current,[18,20,21] otherwise there would be no superconductivity. What remains after this compensation is a part of the superconducting current. The resulting photon spectrum is thereby determined by the inverse of the London penetration depth (due to the Higgs–Meissner effect[18]), but not by the plasma frequency, as in the normal system.

In convential superconductors, which contain positively charged ions, paired electrons, and normal electrons at $T \neq 0$, the photon spectrum is determined by the relation between the vector potential $\mathbf{A}$ and the current $\mathbf{j}$, which is proportional to $\mathbf{A}$; see Eqs. (96.24) and (97.4) of Ref. 20. The analogy with the present case is straightforward. From the

latter equation, for sufficiently low photon momenta we immediately obtain the relation $4\pi\mathbf{j}\simeq -m_\gamma^2(T)\mathbf{A}$ between the Fourier components of the current and the vector potential, where the effective photon mass is

$$m_\gamma(T)\simeq \sqrt{\frac{4\pi e^2\rho_p^*(T)}{m_p^*}}, \quad T<T_{c,p}. \tag{2}$$

Here m_p^* denotes the effective in-medium proton mass, and $\rho_p^*(T)=\rho_p(T_{c,p}-T)/T_{c,p}$ denotes the paired proton density. The choice of a linear temperature dependence for ρ_p^* corresponds to the Ginzburg–Landau approach. A small complex contribution $\sim e^2 f(\omega,k)\exp(-\Delta_p/T)\mathbf{A}$, where $f(\omega,\mathbf{k})$ is a function of the photon frequency ω and momentum $\mathbf{k}$, has been neglected in the above relation between $\mathbf{j}$ and $\mathbf{A}$. More realistically, for T near $T_{c,p}$, one must take into account this off-shell effect for the photon. At lower temperatures, correction terms are exponentially suppressed. Below we take the photon spectrum to be

$$\omega=\sqrt{\mathbf{k}^2+m_\gamma^2}, \tag{3}$$

thus neglecting the aforementioned small polarization effects.

Note that external photons cannot penetrate far into the superconducting region. The photons that we deal with are thermal photons with foregoing dispersion law, governed by the corresponding Bose distribution. In considering neutrino reactions below, we integrate over the photon phase-space volume, thus accurately accounting for the distribution of these photons in warm neutron star matter.

To illustrate more transparently the most important facets of the reconstruction of the photon spectrum in the superconducting region, we consider a two-component, locally neutral system consisting of charged fermions (i.e., the normal subsystem) described by the Dirac field ψ, and a charged condensate (i.e., the superconducting subsystem) described by a condensate wave function

$$\varphi=\varphi_c e^{i\Phi}. \tag{4}$$

The real quantity φ_c is the order parameter of the system, i.e., $\varphi_c^2\sim n_c$, where n_c is the number density of particles in the condensate, and the real value Φ is a phase. In a fermionic system with pairing, the density n_c is proportional to the pairing gap Δ.

The equation for the electromagnetic field A_μ in such a system reads

$$\Box A_\mu=4\pi j_\mu, \tag{5}$$

where the current is

$$j_\mu=ei\bar{\psi}\gamma_\mu\psi-ei(\varphi^*\partial_\mu\varphi-\varphi\partial_\mu\varphi^*)-2e^2|\varphi|^2A_\mu. \tag{6}$$

Substituting Eq. (4) into Eq. (6), we obtain for the electromagnetic current

$$j_\mu=j_\mu^A+\delta j_\mu. \tag{7}$$

where the first term $j_\mu^A=-2e^2\varphi_c^2A_\mu$ is the superconducting current, and the second term δj_μ contains the normal current $j_\mu^{\rm nor}$ and some response $j_\mu^{\rm res}$ from the charged condensate, i.e.,

$$\delta j_\mu=j_\mu^{\rm nor}+j_\mu^{\rm res}=ei\bar{\psi}\gamma_\mu\psi+2e\varphi_c^2\partial_\mu\Phi_0. \tag{8}$$

Due to gauge invariance, the phase $\Phi=\Phi_0+\Phi'$ is not constrained, and Φ_0 can be chosen in such a way that it cancels the normal current, i.e., $\delta j_\mu=0$; otherwise the remaining part of the normal current would destroy superconductivity and the ground state energy would increase. This compensation of the normal current $j_\mu^{\rm nor}$, which in metals and in normal plasma is proportional to the electric field $\mathbf{E}$, is a necessary condition for the existence of superconductivity. Only a diamagnetic part of the fermionic current proportional to the electromagnetic field A_μ may remain. The latter may lead only to a minor ($\sim e^2$) contribution to the unit values of dielectric and diamagnetic constants. The remaining part of the phase Φ' is hidden in the gauge field, resulting in the disappearance of the Goldstone field (see the analogous discussion of the Higgs effect, e.g., in Ref. 22). The total number of degrees of freedom does not change, so the disappearance of the Goldstone field is compensated by the appearance of an extra (third) polarization of the photon. As a result of Eqs. (5) and (7), the electromagnetic field obeys the equation

$$\Box A_\mu=-8\pi e^2\varphi_c^2A_\mu, \tag{9}$$

which immediately yields the photon spectrum in the form (3), where the photon mass is now given by

$$m_\gamma=\sqrt{8\pi e^2\varphi_c^2}. \tag{10}$$

What we have demonstrated is known as the Higgs-Meissner effect: in the presence of a superconducting component, the photon acquires finite mass. We see that in a two-component (normal+superconducting) system, the photon is described by the dispersion relation (3), as it would be in a purely superconducting system, and not by a plasma-like dispersion law, as in the absence of superconductivity. Another way to arrive at Eq. (3) is given in the Appendix in a noncovariant formulation. Similar derivations for different specific physical systems, guided by the general principle of the compensation of the normal currents in a superconductor, can be found in Refs. 18, 20, 21, and 23.

Expressing the amplitude of the condensate field in terms of the paired proton density,[18] one obtains from Eq. (10) the result (2). Taking $m_p^*(\rho_0)\simeq 0.8m_N$ (with m_N the free nucleon mass), with Eqs. (1) and (2) we estimate

$$m_\gamma(\rho=\rho_0,T)[\mathrm{MeV}]\simeq 1.6\sqrt{\frac{T_{c,p}-T}{T_{c,p}}}\ll\omega_{pl}(\rho\sim\rho_0).$$

Due to the rather low effective photon mass in superconducting neutron star matter at $T<T_{c,p}<\omega_{pl}$, one may expect a corresponding increase in the contribution of the above diagrams to neutrino emissivity.

To avoid misunderstanding, we note the following. At the first glance one might suggest that the photon self-energy is completely determined by the above neutrino production diagrams, but with neutrino legs replaced by a photon line. If so, the contributions of the electron-loop and proton-loop diagrams would accurately determine the plasmon spectrum of photon excitations with energy gap equal to a high plasma frequency (at least if one drops small terms proportional to

the proton gap in the calculation of the proton–proton-hole diagram, now with an incoming and outgoing photon, as suggested for the corresponding neutrino process). How does this relate to the massive photon spectrum of superconducting systems? The answer is that in a system with a charged condensate, in addition to the cited photon propagation diagrams, there appear specific diagrams for photon rescattering off the condensate given by terms proportional to $e^2\varphi_c^2 A_\mu A^\mu$ and $2e\varphi_c^2\partial_\mu \Phi A^\mu$ in the corresponding Lagrangian. Their contributions to the equation of motion for the electromagnetic field are, respectively, the last two condensate terms in the electromagnetic current in Eq. (6). The specific condensate diagrams responsible for the compensation of the loop diagram contributions in the photon propagator make no contribution to neutrino emissivity. Indeed, the neutrino legs cannot be directly connected to the photon line via such interactions (without invoking the internal structure of the condensate order parameter φ_c; this contribution is obviously small compared to what we have taken into account). Thus, we have argued that in the presence of superconducting protons, neutrino pairs can be produced in the reaction shown by the above diagrams, where the photons possess rather small masses generated by the Higgs–Meissner mechanism.

Having clarified this important issue, we are ready to calculate the contribution of these processes to neutrino emissivity and compare the result with known emission rates.

3. CALCULATION OF EMISSIVITY

The matrix element of the above diagrams for the ith neutrino species ($i=\{\nu_e,\nu_\mu,\nu_\tau\}$) is

$$\mathscr{M}^{(i)a}=-i\sqrt{4\pi}\,e\,\frac{G}{2\sqrt{2}}\,\varepsilon_\mu^a(\Gamma_\gamma T_p^{(i)\mu\rho}-T_e^{(i)\mu\rho})l_\rho, \qquad (11)$$

where

$$T_j^{(i)\mu\rho}=-\mathrm{Tr}\int\frac{d^4p}{(2\pi)^4}\gamma^\mu i\hat{G}_j(p)W_j^{(i)\rho}\,i\hat{G}_j(p+k),$$

$$j=\{e,p\}, \qquad (12)$$

and

$$\hat{G}_j(p)=(\hat{p}+m_j)\left\{\frac{1}{p^2-m_j^2}+2\pi i\,n_j(p)\,\delta(p^2-m_j^2)\,\theta(p_0)\right\} \qquad (13)$$

is the in-medium electron (proton) Green's function; $n_j(p)=\theta(p_{Fj}-p)$; ε_μ^a is the corresponding polarization four-vector of the massive photon, with three polarization states in superconducting matter. The factor Γ_γ takes into account nucleon-nucleon correlations in the photon vertex. The quantity $G=1.17\cdot10^{-5}$ GeV^{-2} is the Fermi constant of the weak interaction. Above, l_p denotes the neutrino weak current. The electron and proton weak currents are

$$W_e^{(i)\rho}=\gamma^\rho(c_V^{(i)}-c_A^{(i)}\gamma_5),\quad W_p^\rho=\gamma^\rho(\kappa_{pp}-g_A\gamma_{pp}\gamma_5), \qquad (14)$$

where $c_V^{(\nu_e)}=c_V^{(+)}=1+4\sin^2\vartheta_W\simeq1.92$ and $c_V^{(\nu_\mu)}=c_V^{(\nu_\tau)}=c_V^{(-)}=1-4\sin^2\vartheta_W\simeq0.08$; ϑ_W is the Weinberg angle, and $c_A^{(\nu_e)}=-c_A^{(\nu_\mu,\nu_\tau)}=1$. Proton coupling is corrected by nucleon-nucleon correlations, i.e., by the factors κ_{pp} and γ_{pp}.[24]

Integrating Eq. (12) over the energy variable, we obtain for the ith neutrino species

$$-i(T_p^{(i)\mu\rho}-T_e^{(i)\mu\rho})=\tau_t^{(i)}P^{\mu\rho}+\tau_l^{(i)}F^{\mu\rho}+\tau_5^{(i)}P_5^{\mu\rho}, \qquad (15)$$

$$P^{\mu\rho}=\left(g^{\mu\rho}-\frac{k^\mu k^\rho}{k^2}+F^{\mu\rho}\right),\quad F^{\mu\rho}=\frac{j^\mu j^\rho}{k^2[(k\cdot u)^2-k^2]},$$

$$P_5^{\mu\rho}=\frac{i}{\sqrt{k^2}}\,\varepsilon^{\mu\rho\delta\lambda}k_\delta u_\lambda, \qquad (16)$$

where $j^\mu=(k\cdot u)k^\mu-u^\mu k^2$, $(k\cdot u)=k_\mu u^\mu$, $k^\mu=(\omega,\mathbf{k})$, $k^2=k_\mu k^\mu=\omega^2-\mathbf{k}^2$. The four-velocity u^μ of the medium is introduced for the sake of covariant notation. The transverse (τ_t), longitudinal (τ_l), and axial (τ_5) components of the tensors in Eq. (15) yield

$$\tau_t^{(i)}=\tau_{te}^{(i)}-\tau_{tp}^{(i)}=2c_V^{(i)}(A_e+k^2B_e)-2c_V^{(-)}R_\kappa(A_p+k^2B_p), \qquad (17)$$

$$\tau_l^{(i)}=\tau_{le}^{(i)}-\tau_{lp}^{(i)}=4k^2[c_V^{(i)}B_e-c_V^{(-)}R_\kappa B_p], \qquad (18)$$

$$\tau_5^{(i)}=\tau_{5e}^{(i)}-\tau_{5p}^{(i)}=(k^2)^{3/2}[c_A^{(i)}C_e-g_A\gamma_{pp}C_p], \qquad (19)$$

where $R_\kappa=\kappa_{pp}/c_V^{(-)}$, and

$$A_j=\int\frac{d^3p}{(2\pi)^3}\frac{n_j(p)}{E_p^{(j)}}+\frac{k^2}{2}\left(1+\frac{k^2}{2m_j^2}\right)m_jC_j, \qquad (20)$$

$$B_j=\int\frac{d^3p}{(2\pi)^3}\frac{n_j(p)}{2E_p^{(j)}}\frac{1-(\mathbf{pk})^2/E_p^{(j)2}\mathbf{k}^2}{(\omega-\mathbf{pk}/E_p^{(j)})^2-k^4/4E_p^{(j)4}}, \qquad (21)$$

$$C_j=\int\frac{d^3p}{(2\pi)^3}\,n_j(p)\,\frac{m_j}{E_p^{(j)3}}\left[\left(\omega-\frac{\mathbf{pk}}{E_p^j}\right)^2-\frac{k^4}{4E_p^{(j)4}}\right]^{-1},$$

$$E_p^{(j)}=\sqrt{m_j^2+\mathbf{p}^2}. \qquad (22)$$

Here we note that the contribution of the axial component τ_5 to the resulting neutrino emissivity is small ($\tau_5/\tau_t\sim m_\gamma^2\tau_5/\omega^2\tau_l\sim m_\gamma/m_N^*$ for protons and $\sim(m_\gamma m_e/p_{Fe}^2)\times\ln(p_{Fe}/m_e)$ for electrons), so that it will be omitted.

The squared matrix element (11) for a certain neutrino species, summed over the lepton spins and averaged over the three photon polarizations, can be cast in the form

$$\overline{\sum|\mathscr{M}^{(i)}|^2}=\frac{4}{3}\,\pi e^2G^2\left[\tau_t^{(i)2}\left(2\omega_1\omega_2+2\frac{(\mathbf{kq}_1)(\mathbf{kq}_2)}{\mathbf{k}^2}\right)-\tau_l^{(i)2}\left(\omega_1\omega_2+\mathbf{q}_1\mathbf{q}_2-2\frac{(k\cdot q_1)(k\cdot q_2)}{k^2}-2\frac{(\mathbf{kq}_1)(\mathbf{kq}_2)}{\mathbf{k}^2}\right)\right], \qquad (23)$$

where $(k\cdot q_{1,2})=\omega\omega_{1,2}-(\mathbf{kq}_{1,2})$, and $\omega_{1,2}$ and $\mathbf{q}_{1,2}$ denote the frequencies and momenta of the neutrino and antineutrino. We have also used the fact that $\mathrm{Tr}\{l^\mu l^\nu\}=8[q_1^\mu q_2^\nu+q_2^\mu q_1^\nu-g^{\mu\nu}(q_1\cdot q_2)-4i\varepsilon^{\mu\nu\lambda\rho}q_{1\lambda}q_{2\rho}]$.

The emissivity of our processes is given by

$$\varepsilon_\nu^\gamma = \int \frac{d^3k}{(2\pi)^3 2\omega}\frac{d^3q_1}{(2\pi)^3 2\omega_1}\frac{d^3q_2}{(2\pi)^3 2\omega_2} \times \frac{\omega_1+\omega_2}{\exp[(\omega_1+\omega_2)/T]-1} \times \sum_{i=\nu_e,\nu_\mu,\nu_\tau} \overline{\sum |\mathscr{M}^{(i)}|^2}(2\pi)^4\delta^4(k-q_1-q_2). \quad (24)$$

Substituting Eq. (23) into Eq. (24), we finally obtain

$$\varepsilon_\nu^\gamma = \frac{T^5}{9(2\pi)^3}\,\pi e^2 G^2 \alpha^2 I,$$

$$I = \int_\alpha^\infty \frac{d\xi\xi}{e^\xi - 1}\sqrt{\xi^2-\alpha^2}\left[\tau_t^2\left(\frac{\alpha^2}{\xi^2}\right)+\tau_l^2\left(\frac{\alpha^2}{\xi^2}\right)\right], \quad (25)$$

where $\alpha = m_\gamma/T$, and

$$\tau_t^2(x) \approx 4\sum_{i=\nu_e,\nu_\mu,\nu_\tau}\left[c_V^{(-)}R_\kappa\frac{\rho_p}{2m_p^*}(1+x)-c_V^{(i)} \times\left(\frac{3}{8\pi}\rho_p\right)^{2/3}\left(1+\frac{x}{2}\right)\right]^2, \quad (26)$$

$$\tau_l^2(x) \approx 4x^2\sum_{i=\nu_e,\nu_\mu,\nu_\tau}\left[c_V^{(-)}R_\kappa\frac{\rho_p}{2m_p^*}-c_V^{(i)}\left(\frac{3}{8\pi}\rho_p\right)^{2/3}\right]^2. \quad (27)$$

Some numerically small terms have been dropped in Eq. (26).

The integral I in Eq. (25) can be calculated analytically in the two limiting cases, $\alpha \ll 1$ and $\alpha \gg 1$:

$$I(\alpha \gg 1) \approx \frac{\sqrt{2\pi}}{2}\,\alpha^{3/2}\left(1+\frac{3}{2\alpha}\right)e^{-\alpha}[\tau_l^{(2)}(1)+\tau_t^2(1)], \quad (28)$$

$$I(\alpha \ll 1) \approx 2\zeta(3)[\tau_l(0)+\tau_t^2(0)], \quad \zeta(3)\simeq 1.202. \quad (29)$$

Thus, combining Eqs. (1) and (25)–(28), we obtain an estimate for emissivity of our reactions (we present here the result for $m_\gamma > T$ and for three neutrino species):

$$\varepsilon_\nu^\gamma\left[\frac{\text{erg}}{\text{cm}^3\cdot\text{s}}\right] \approx 2.6\cdot 10^{25} T_9^{3/2}\exp\left(-\frac{m_\gamma}{T}\right)\left(\frac{m_\gamma}{\text{MeV}}\right)^{7/2} \times\left(\frac{\rho}{\rho_0}\right)^{8/3}\left(1+\frac{3}{2}\frac{T}{m_\gamma}\right)[1+\eta], \quad (30)$$

$$\eta = 0.0003\,R_\kappa^2\left(\frac{m_p}{m_p^*}\right)^2\left(\frac{\rho}{\rho_0}\right)^{4/3}-0.035\,R_\kappa\frac{m_p}{m_p^*}\left(\frac{\rho}{\rho_0}\right)^{2/3}. \quad (31)$$

Here T_9 denotes temperature measured in units of 10^9 K. The unity in square brackets in Eq. (30) corresponds to the electron–electron-hole diagram, whereas the factor η is related to the proton–proton-hole (first term in Eq. (31)) and the interference diagrams (second term in Eq. (31)).

Emissivity given by Eq. (30) varies with temperature as $T^{3/2}\exp(-m_\gamma/T)$, whereas emissivity of the modified URCA process varies as $T^8\exp[-(\Delta_p+\Delta_n)/T]$ in the region of proton ($\Delta_p \neq 0$) and neutron ($\Delta_n \neq 0$) pairing. Hence, one can expect that the process $\gamma_m \to \nu\bar{\nu}$ will dominate at comparatively low temperatures, when $\Delta_p(T)+\Delta_n(T)-m_\gamma(T)>0$ and $T<T_{c,p}$.

4. NUMERICAL ESTIMATES

To obtain quantitative estimates we need the values of the nucleon-nucleon correlation factors κ_{pp} and Γ_γ. According to Ref. 24, we can exploit

$$\kappa_{pp} = c_V^{(-)} - 2\,f_{np}C_0A_{nn}\Gamma(f_{nn}), \quad (32)$$

where $f_{np}\simeq -0.75$ and $f_{nn}\simeq 1.25$ are the constants in the theory of finite Fermi systems;[19,24] $C_0^{-1} = m_n^* p_{Fn}/\pi^2$ is the density of states at the Fermi surface; A_{nn} is the neutron–neutron-hole loop,

$$C_0A_{nn} = iC_0\int\frac{d^3p}{(2\pi)^4}\,G_n(p+k)G_n(p) \approx \frac{p_{Fn}^2k^2}{6m_n^*\omega^2}, \quad (33)$$

for values of $\omega \gg |\mathbf{k}|p_{Fn}/m_n^*$ of interest, and $\Gamma^{-1}(f_{nn}) = 1 - 2\,f_{nn}C_0A_{nn}$.

We note that the second term in Eq. (32) is not proportional to a small factor $c_V^{(-)}$, because the nucleon-nucleon correlations also allow for emission of $\nu\bar{\nu}$-pairs from the nn^{-1} loop. Numerical estimates of the ratio R_κ are as follows: for $\alpha \gg 1$, we have $R_\kappa \simeq 1.6$ for $\rho = \rho_0$, $m_n^*(\rho_0) \simeq 0.8m_n$, and $R_\kappa \simeq 2.1$ for $\rho = 2\rho_0$, $m_n^*(2\rho_0) \simeq 0.7m_n$; for $\alpha \ll 1$, we obtain $R_\kappa \simeq 1$ and correlation effects are negligible. The in-medium renormalization of the proton electric charge included in the factor Γ_γ can be also expressed in terms of the constants in the theory of finite Fermi systems and the proton–proton loop factor (A_{pp}); see Ref. 19. The latter is suppressed at relatively low proton densities. We can therefore take $\Gamma_\gamma \approx 1$. With these estimates, we observe that the main contribution to neutrino emissivity comes from electron–electron-hole processes.

The ratio of emissivity ε_ν^γ (30) to emissivity $\varepsilon_\nu^{\text{FM}}$ of the modified URCA process, $R_{\text{FM}} = \varepsilon_\nu^\gamma/\varepsilon_\nu^{\text{FM}}$, is

$$R_{\text{FM}} \approx 15\cdot 10^4 T_9^{-13/2}\exp\left(\frac{\Delta_n+\Delta_p-m_\gamma}{T}\right)\left(\frac{m_\gamma}{\text{MeV}}\right)^{7/2} \times\left(1+\frac{3}{2}\frac{T}{m_\gamma}\right)\left(\frac{\rho}{\rho_0}\right)^2\frac{m_n^3m_p}{m_n^{*3}m_p^*}[1+\eta]. \quad (34)$$

For further estimates we need the values of the neutron and proton gaps, which are unfortunately model-dependent. For instance, the evaluation in Ref. 25 yields $\Delta_n(0) \simeq 8.4T_{c,n} \simeq 0.6$ MeV, $T_{c,n} \simeq 0.07$ MeV for $^3P_2$ neutron pairing at $\rho = \rho_0$, and $\Delta_p(0) \simeq 1.76T_{c,p} \simeq 3$ MeV, $T_{c,p} \simeq 1.7$ MeV for $1S$ proton pairing, while Ref. 26 uses $\Delta_n(0) \simeq 2.1$ MeV, $T_{c,n} \simeq 0.25$ MeV and $\Delta_p(0) \simeq 0.7$ MeV, $T_{c,p} \simeq 0.4$ MeV for $\rho = \rho_0$. Employing these estimates of the zero-temperature gaps, its temperature dependence, and the photon effective mass, we obtain from Eq. (34) the temperature dependence of the ratio R_{FM}.

In order to find the lower temperature limit at which the processes $\gamma_m \to \nu\bar{\nu}$ are still operative, we need to compare the value ε_ν^γ with photon emissivity at the neutron star surface, $\varepsilon_\gamma^s = 3\sigma T_s^4/R$, where σ is the Stefan–Boltzmann con-

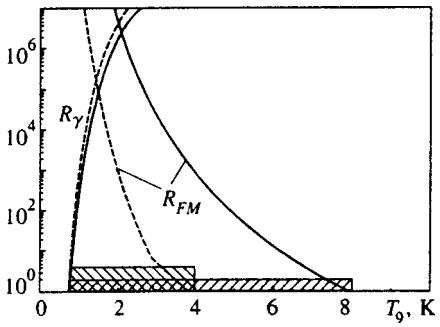


FIG. 1. Temperature dependence of the ratios $R_{\rm FM}$ and R_γ at nucleon density $\rho=\rho_0$. Solid curves correspond to the parameter choice of Ref. 25, whereas the dashed curves depict results with parameters of Ref. 26. Shaded bars indicate the temperature regions in which cooling via massive photon decay is more efficient than standard cooling processes.

stant, T_s denotes the surface temperature of the star, and R is the star's radius. By employing a relation[27] between the surface and interior temperatures, we obtain for $R_\gamma=\varepsilon_\nu^\gamma/\varepsilon_\gamma^s$

$$R_\gamma \approx 1.2\cdot 10^9 T_9^{-0.7}\exp\left(-\frac{m_\gamma}{T}\right)\left(\frac{m_\gamma}{\rm MeV}\right)^{7/2} \times\left(1+\frac{3}{2}\frac{T}{m_\gamma}\right)\left(\frac{\overline{\rho}}{\rho_0}\right)^{8/3}[1+\eta], \tag{35}$$

where the star radius and mass are taken to be 10 km and $1.4M_\odot$, with $M_\odot$ the solar mass and $\overline{\rho}$ some averaged value of the density in the neutron star interior.

The ratios $R_{\rm FM}$ and R_γ are plotted as a function of the temperature in Fig. 1 for both of the foregoing parameter choices. We see that our new processes are operative in the temperature range $1\cdot 10^9$ K$\lesssim T\lesssim 8\cdot 10^9$ K for the parameter choice of Ref. 25, and $1\cdot 10^9$ K$\lesssim T\lesssim 4\cdot 10^9$ K for the parameters of Ref. 26. As one observes in Fig. 1, within these intervals the new cooling channel might exceed known cooling processes by up to a factor 10^6.

5. CONCLUDING REMARKS

As mentioned above, for $T>T_{c,n}$, $T_{c,p}$, i.e., in a normal plasma region of the star crust and star interior, photons with approximately the electron plasma frequency[2)] ω_{pl} can decay into neutrino pairs, as has been shown in previous estimates.[10] At $T<T_{c,p}$, however, we are already dealing with massive photons in the region of proton pairing, and our new reaction channels can significantly contribute to cooling.

Our processes can also occur in a charged-pion (or kaon) condensate state but they are suppressed due to the high effective photon mass[3)] $m_\gamma\simeq\sqrt{8\pi e^2\varphi_c^2}\simeq 6$ MeV for the condensate field $\varphi_c\simeq 0.1 m_\pi\simeq 14$ MeV.

In deriving the value of $\varepsilon_\nu^{\rm FM}$ used above, one describes the nucleon-nucleon interaction essentially by free one-pion exchange. In reality, however, at $\rho>(0.5-1)\rho_0$ the total nucleon-nucleon interaction does not reduce to free one-pion exchange, because of the strong polarization of the medium, whereby a significant part comes from in-medium pionic excitations.[5–7,24] Occurring in intermediate states of the reaction, the in-medium pions can also decay into $e\overline{\nu}$, or first into a nucleon–nucleon-hole, which then radiates $e\overline{\nu}$, thereby substantially increasing the resulting emissivity. Other reaction channels such as $n\to n_{\rm pair}\nu\overline{\nu}$ and $p\to p_{\rm pair}\nu\overline{\nu}$ open up in the superfluid phase with paired nucleons,[6,24,28] where $n_{\rm pair}$ ($p_{\rm pair}$) means a paired neutron (proton). All these reaction channels give rise to a larger contribution to emissivity than that of the modified URCA process estimated via free one-pion exchange. Above we compared ε_ν^γ with $\varepsilon_\nu^{\rm FM}$ just because the latter is used in the standard scenarios of neutron star cooling.

As we also mentioned in the Introduction, there are other processes like those considered above. Emissivity of the process $p\gamma_m\to p_{\rm pair}\nu\overline{\nu}$ is substantially suppressed (at least by a factor e^2 and also due to a much smaller phase-space volume) compared to that of the process $p\to p_{\rm pair}\nu\overline{\nu}$. According to simple estimates, e.g., using Eq. (22) of Ref. 16, the process $e\gamma\to e\nu\overline{\nu}$ makes a very small contribution to emissivity both in the inner crust and in the interior of neutron stars, even when one neglects the photon mass. Thus we may conclude that the process $e\gamma_m\to e\nu\overline{\nu}$ also leads to a minor contribution to emissivity at the densities and temperatures under consideration.

In summary, the processes $\gamma_m\to ee^{-1}+pp^{-1}\to\nu\overline{\nu}$ might be operative over some temperature interval $T\simeq 10^9-10^{10}$ K, $T<T_{c,p}$, and together with other in-medium modified processes,[8] they should be incorporated into computer simulations of neutron star cooling.

We acknowledge V. M. Osadchiev for fruitful discussions. The research described in this publication was made possible in part by Grants N3W000 from the International Science Foundation and N3W300 from the International Science Foundation and the Russian Government. B. K. and E. E. K. are supported by BMBF Grant 06DR666. E. E. K. acknowledges the support of the Heisenberg–Landau program.

APPENDIX

We can also achieve the same results that led to Eq. (10) by starting with Maxwell's equations (in obvious notation):

$$i\mathbf{k}\cdot\mathbf{E}=4\pi\tilde{\rho},\quad i\mathbf{k}\times\mathbf{B}=4\pi\mathbf{j}-i\omega\mathbf{E},$$

$$\mathbf{k}\cdot\mathbf{B}=0,\quad \mathbf{k}\times\mathbf{E}=\omega\mathbf{B},$$

where the charge density $\overline{\rho}$ is the superposition of the density of free charges and the density of bound charge. Full free charge density being zero in our case due to local electroneutrality. The current $\mathbf{j}$ is a superposition of an external test current and the induced current:

$$\mathbf{j}=\mathbf{j}^{\rm ext}+\mathbf{j}^{\rm ind}.$$

In normal systems, the induced current (i.e., the current of nonpaired charged particles) $\mathbf{j}^{\rm ind}=\mathbf{j}^{\rm nor}$ is related to $\mathbf{E}$ via longitudinal ϵ_l and transverse ϵ_t dielectric constants. This connection results in longitudinal and transverse branches of the electromagnetic excitations, with an effective photon gap equal to the plasma frequency ω_{pl}.[10] In contrast, in a superconducting system the condensate makes two other contributions to the current, namely $\mathbf{j}^A=-2e^2\varphi_c^2\mathbf{A}$ and $\mathbf{j}^{\rm res}$

$=2e\varphi_c^2\nabla\Phi$. Letting $\Phi=\Phi_{(1)}+\Phi_{(2)}$, we have $\mathbf{j}^{\rm res}=\mathbf{j}^{\rm res}_{(1)}+\mathbf{j}^{\rm res}_{(2)}$. These two terms are determined as follows. As we have argued above, superconductivity requires the compensation of the normal component of the current proportional to **E**, i.e., we can take $\mathbf{j}^{\rm nor}+\mathbf{j}^{\rm res}_{(1)}=0$. Only small contributions $\sim e^2\exp(-\Delta_p/T)\omega^2\mathbf{A}$ and $\sim e^2\exp(-\Delta_p/T)\mathbf{k}^2\mathbf{A}$, as well as a small imaginary contribution $\sim ie^2F(\omega,\mathbf{k})\exp(-\Delta_p/T)\mathbf{A}$, where F is some function of ω and **k**, can still remain from the value $\mathbf{j}^{\rm nor}$ (see Eqs. (96.24) and (97.4) of Ref. 20). We neglect these small contributions. The part of the current $\sim\nabla\Phi_{(2)}$ can be hidden in $\mathbf{j}^A$ by a gauge transformation of the field **A**. We then have

$$i\mathbf{k}\times\mathbf{B}\simeq\mathbf{j}^A-i\omega\mathbf{E}.$$

Taking the vector product of this equation with **k**, we obtain

$$(\omega^2-\mathbf{k}^2-8\pi e^2\varphi_c^2)\mathbf{B}=0.$$

From this relation we observe that the electromagnetic excitations possess the mass given by Eq. (10). Hence, we have demonstrated that one can obtain the well-known plasma photon spectrum for a normal system, and at the same time one can obtain a massive photon spectrum and the Higgs–Meissner effect in a system with a charged condensate.

*)E-mail: voskre@rzri6f.gsi.de
†)E-mail: kolomei@tpri6f.gsi.de

1)Note that in conventional nuclear physics one usually employs particle–hole diagrams even at zero temperature, thereby considering nuclear matter to be normal. Small effects of pairing can be neglected, since the typical energy in a nucleonic particle–hole diagram is of the order of the Fermi energy ϵ_F, and $\epsilon_F\gg\Delta$ holds.[7,18,19]

2)A rather small extra contribution also comes from the proton-proton-hole diagram.

3)For simplicity, in this estimate the peculiarities of a condensate with nonvanishing momentum[7] are ignored.

[1] S. Shapiro and S. A. Teukolsky, *Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects*, Wiley, New York (1983), Ch. 1; *Neutron Stars*, D. Pines, R. Tamagaki, and S. Tsuruta (eds.), Addison-Weseley, New York (1992); H. Umeda, K. Nomoto, S. Tsuruta *et al.*, Astrophys. J. **431**, 123 (1994); D. Page, E-print archive, astro-ph/9706259.
[2] J. N. Bahcall and R. A. Wolf, Phys. Rev. B **140**, 1445 (1965); S. Tsuruta and A. G. W. Cameron, Can. J. Phys. **43**, 2056 (1965).
[3] B. Friman and O. V. Maxwell, Astrophys. J. **232**, 541 (1979); O. V. Maxwell, Astrophys. J. **231**, 201 (1979).
[4] S. Tsuruta, Phys. Rep. **56**, 237 (1979); K. Nomoto and S. Tsuruta, Astrophys. J. Lett. **250**, 19 (1981); Ch. Schaab, F. Weber, M. K. Weigel, and N. K. Glendenning, Nucl. Phys. A **605**, 531 (1996).
[5] D. N. Voskresensky and A. V. Senatorov, JETP Lett. **40**, 1212 (1984); D. N. Voskresensky and A. V. Senatorov, Zh. Éksp. Teor. Fiz. **90**, 1505 (1986) [Sov. Phys. JETP **63**, 885 (1986)].
[6] A. V. Senatorov and D. N. Voskresensky, Phys. Lett. B **184**, 119 (1987).
[7] A. B. Migdal, E. E. Saperstein, M. A. Troitsky, and D. N. Voskresensky, Phys. Rep. **192**, 179 (1990).
[8] Ch. Schaab, D. Voskresensky, A. D. Sedrakian, F. Weber, and M. K. Weigel, Astron. Astrophys. **321**, 591 (1997).
[9] G. Flowers, P. G. Sutherland, and J. R. Bond, Phys. Rev. D **12**, 315 (1975).
[10] J. B. Adams, M. A. Ruderman, and C.-H. Woo, Phys. Rev. **129**, 1383 (1963); D. A. Dicus, Phys. Rev. D **6**, 941 (1972); V. N. Oraevskii, V. B. Semikoz, and Ya. A. Smorodinsky, Part. and Nuclei **55**, 312 (1995); J. C. D'Olivio, J. F. Nieves, and P. Pal, Phys. Rev. D **40**, 3679 (1989).
[11] M. Soyeur and G. E. Brown, Nucl. Phys. A **324**, 464 (1979).
[12] B. Pontecorvo, Zh. Éksp. Teor. Fiz. **36**, 1615 (1959) [Sov. Phys. JETP **9**, 1148 (1959)].
[13] G. G. Festa and M. A. Ruderman, Phys. Rev. **180**, 1227 (1969); D. A. Dicus, E. W. Kolb, D. N. Schramm, and D. L. Tubbs, Astrophys. J. **210**, 481 (1976).
[14] H. J. Chiu and P. Morrison, Phys. Rev. Lett. **5**, 573 (1960); H. J. Chiu, Phys. Rev. **123**, 1040 (1961).
[15] H. J. Chiu and R. C. Stabler, Phys. Rev. **122**, 1317 (1961).
[16] V. I. Ritus, Zh. Éksp. Teor. Fiz. **41**, 1285 (1961) [Sov. Phys. JETP **14**, 915 (1962)].
[17] V. Petrosian, G. Beaudet, and E. E. Salpeter, Phys. Rev. **154**, 1445 (1967); D. A. Dicus, Phys. Rev. D **6**, 941 (1972).
[18] E. M. Lifshitz and L. P. Pitaevskii, *Statisticheskaya Fizika*, Chast' II, Nauka, Moscow (1978) [*Statistical Physics*, Part 2, Pergamon Press, New York (1980)].
[19] A. B. Migdal, *Theory of Finite Systems and Properties of Atomic Nuclei* [in Russian], Nauka, Moscow (1983).
[20] E. M. Lifshitz and L. P. Pitaevskii, *Fizicheskaya Kinetika*, Nauka, Moscow (1979) [Physical Kinetics, Pergamon Press, New York (1981)].
[21] S. J. Putterman, *Superfluid Hydrodynamics*, North-Holland, Amsterdam (1974); L. D. Landau and E. M. Lifshitz, *Gidrodinamika*, Nauka, Moscow (1986) [*Fluid Mechanics*, Pergamon Press, New York (1987)].
[22] V. B. Berestetskii, *Problems of Elementary Particle Physics* [in Russian], Nauka, Moscow (1979).
[23] B. J. Harrington and H. K. Shepard, Phys. Rev. D **19**, 1713 (1979); D. N. Voskresensky and N. Yu. Anisimov, Zh. Éksp. Teor. Fiz. **78**, 28 (1980) [Sov. Phys. JETP **51**, 13 (1980)]; D. N. Voskresensky, Zh. Éksp. Teor. Fiz. **105**, 1473 (1994) [JETP **78**, 793 (1994)].
[24] D. N. Voskresensky and A. V. Senatorov, Yad. Fiz. **45**, 657 (1987) [Sov. J. Nucl. Phys. **45**, 411 (1987)].
[25] T. Takatsuka, Prog. Theor. Phys. **48**, 1517 (1972).
[26] D. Pines and M. A. Alpar, Nature (London) **316**, 27 (1985).
[27] E. H. Gudmundsson, C. J. Pethick, and R. I. Epstein, Astrophys. J. **272**, 286 (1983).
[28] E. Flowers, M. Ruderman, and P. Sutherland, Astrophys. J. **205**, 541 (1976).

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

Variational estimates of the diffusion coefficient of cosmic rays in a random magnetic field

V. N. Zirakashvili*)

Institute of Geomagnetism, the Ionosphere, and Radiowave Propagation, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 2 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 398–405 (August 1998)

A variational method is used to obtain estimates of the effective particle transport coefficients in a random static magnetic field. The particle propagation is described by an anisotropic diffusion equation. The diffusion coefficient parallel to the local magnetic field is much greater than the transverse diffusion coefficient. For large-scale magnetic-field variations the diffusion is described by effective coefficients. The variational approach can be used to find the effective parallel and perpendicular diffusion coefficients. It was shown that the instability growth rate of the magnetic field lines determines the upper estimate of the effective transverse diffusion coefficient. © *1998 American Institute of Physics.* [S1063-7761(98)00208-X]

1. INTRODUCTION

The motion of charged particles in a random magnetic field has frequently been examined in connection with the propagation of cosmic rays in interplanetary and interstellar media.[1–4] Scattering of particles by small-scale magnetic field variations (with scales of the order of the particle gyroradius) causes them to become isotropic and gives rise to diffusion, which is usually strongly anisotropic: the diffusion coefficient parallel to the field is much greater than the perpendicular diffusion coefficient. The presence of a large-scale random magnetic field component enhances the transverse diffusion (known as anomalous diffusion) and makes the overall diffusion more isotropic. To be more specific, we assume that time-dependent diffusion equation

$$\frac{\partial f}{\partial t}=\nabla_i D_{ij}\nabla f, \tag{1}$$

where D_{ij} is the symmetric diffusion tensor. Here we shall consider the diffusion tensor given in the following form:

$$D_{ij}=(D_\parallel-D_\perp)b_i b_j+D_\perp\delta_{ij},$$
$$D_{ij}^{-1}=(D_\parallel^{-1}-D_\perp^{-1})b_i b_j+D_\perp^{-1}\delta_{ij}. \tag{2}$$

Here $D_\parallel$ and $D_\perp$ are the diffusion coefficients parallel and perpendicular to the magnetic field (we shall subsequently assume that these are independent of position), and $\mathbf{b}$ is the unit vector parallel to the magnetic field: $\mathbf{b}=\mathbf{B}/B$. Fluctuations of the magnetic field give rise to fluctuations of the diffusion tensor and the problem involves averaging Eq. (1) over these fluctuations. We shall subsequently consider that the magnetic field is static and we shall analyze a steady-state variant of Eq. (1). This is correct if the diffusion coefficient is so large that over the characteristic time on which the field varies, the particle diffuses over a distance greater than the correlation length of the field (more strictly, the time of variation of the field should exceed the time required by the particle to ''forget'' the field line; see Conclusions). So far, the problem of averaging the diffusion in a static field has only been solved for trivial cases (for example, if the diffusion tensor only depends on a single coordinate), except for the result obtained by Dykhne.[5] An exact result exists for the two-dimensional problem and the diffusion tensor (2) with an isotropic random magnetic field, for which the effective diffusion coefficient has the form

$$D^*=\sqrt{D_\parallel D_\perp}.$$

The case of the average magnetic field is of interest in the theory of cosmic ray propagation. Averaging Eq. (1) using perturbation theory (see Ref. 4) shows that the small parameter of the problem is $\langle A^2\rangle\sqrt{D_\parallel/D_\perp}$. Here the vector amplitude of the random magnetic field is defined as $\mathbf{A}=(\mathbf{B}-\mathbf{B}_0)/B_0$, $B_0=\langle\mathbf{B}\rangle$ and the angular brackets denote averaging over volume. The effective transverse diffusion coefficient may be expressed as the series

$$D_\perp^*=D_\perp\left(1+C_1\langle A^2\rangle\sqrt{\frac{D_\parallel}{D_\perp}}+C_2\langle A^2\rangle^2\frac{D_\parallel}{D_\perp}+\ldots\right), \tag{3}$$

where $C_1, C_2\ldots$ are numerical coefficients. In the case of practical interest $\langle A^2\rangle\leqslant 1$, perturbation theory is unsuitable since $D_\parallel\gg D_\perp$ and the parameter in question is large. Under these conditions, it may be interesting to determine the limits of variation of the effective transverse diffusion coefficient and for this purpose we use a variational approach.

2. VARIATIONAL ESTIMATES OF THE EFFECTIVE PARAMETERS

In order to obtain variational estimates of the effective diffusion coefficient for the equation

$$\nabla_i D_{ij}\nabla f=0 \tag{1′}$$

we shall consider the functionals

$$L^{(1)}=\langle D_{ij}g_i g_j\rangle,\quad L^{(2)}=\langle D_{ij}^{-1}j_i j_j\rangle. \tag{4}$$

Here **g** and **j** are vectors which depend on position. We can show that subject to the additional conditions

$$\mathbf{G}=\langle\mathbf{g}\rangle, \quad \operatorname{curl}\mathbf{g}=0, \quad \mathbf{J}=\langle\mathbf{j}\rangle, \quad \operatorname{div}\mathbf{j}=0 \tag{5}$$

(where **G** and **J** are constant vectors, the first two conditions refer to the first functional, and the third and fourth refer to the second), these functionals have extreme values if

$$g_i=\nabla_i f, \quad j_i=-D_{ij}\nabla_j f, \tag{6}$$

and the function f satisfies Eq. (1′) (see Refs. 6 and 7). In other words, Eq. (1′) may be obtained if the first functional in Eq. (4) is varied subject to the second additional condition (5). These extrema are minima if the components of the diffusion tensor are positive definite quadratic coefficients, and the value of the functionals at the minimum is given by

$$L^{(1)}_{\min}=D^*_{ij}G_iG_j, \quad L^{(2)}_{\min}=(D^*_{ij})^{-1}J_iJ_j. \tag{7}$$

Here the effective diffusion tensor D^*_{ij} gives the relation between the average current and the gradient:

$$J_i=-D^*_{ij}G_j. \tag{8}$$

Equality (7) implies that the following inequalities are satisfied:

$$D^*_{ij}G_iG_j\leqslant\langle D_{ij}g_ig_j\rangle, \quad (D^*_{ij})^{-1}J_iJ_j\leqslant\langle D^{-1}_{ij}j_ij_j\rangle, \tag{9}$$

where g_i and j_i are any functions satisfying the conditions (5). By selecting trial functions, we can use the inequalities (9) to obtain upper and lower constraints on the effective diffusion coefficient. If the trial function depends on some nonrandom function, this function may be obtained by solving the variational problem (the right-hand sides of the inequalities (9) must be minimized). The smaller the right-hand sides of the inequalities (9), the closer are the limits of variation of the diffusion coefficient to the accurate solution of the problem.

When using the inequalities (9), we shall assume that the average over volume is equal to the average over an ensemble of realizations of the random field (for a justification of this assumption and the mathematical aspects of averaging Eq. (1′) see Ref. 8).

3. ESTIMATES OF THE EFFECTIVE PARALLEL DIFFUSION COEFFICIENT

We shall use the following trial functions to obtain estimates of the effective parallel diffusion coefficient:

$$\mathbf{g}=\mathbf{G}, \quad \mathbf{j}=\mathbf{B}J/B_0. \tag{10}$$

We can easily see that the conditions (5) are satisfied (the average gradient **G** and the average current **J** are directed parallel to the average magnetic field). Substituting these trial functions into inequality (9) with the diffusion tensor (2), we obtain

$$D_\parallel B_0^2/\langle B^2\rangle\leqslant D^*_\parallel\leqslant D_\parallel\langle b_\parallel^2\rangle+D_\perp\langle b_\perp^2\rangle. \tag{11}$$

Here $\mathbf{b}_\parallel$ and $\mathbf{b}_\perp$ are the components of the unit vector parallel and perpendicular to the average field.

4. ESTIMATES OF THE EFFECTIVE PERPENDICULAR DIFFUSION COEFFICIENT

If the problem is two-dimensional (the vector of the magnetic field lies in the xz plane and does not depend on y), we take the trial function for the gradient in the following form ($\mathbf{e}_y$ is the unit vector of the y axis):

$$\mathbf{g}=\mathbf{e}_y\times\mathbf{B}\frac{G}{B_0}. \tag{12}$$

In this case, the first pair of conditions (5) is satisfied (the average gradient is directed along the x axis and the average field along the z axis). Substituting the trial function (12) into the first inequality (9) yields the estimate

$$D^*_\perp\leqslant D_\perp\langle B^2\rangle/B_0^2. \tag{13}$$

For the three-dimensional problem we shall confine our analysis to perturbations of the magnetic field perpendicular to the average field directed along the z axis (this situation arises for Alfvén wave turbulence, for example). We take the trial function for the gradient in the form

$$\mathbf{g}=\mathbf{G}+\nabla\int d^3r'd^2r_{1\perp}N(\mathbf{r},\mathbf{r}')G_kA_k \times(z',\mathbf{r}_{1\perp})q(z-z',\mathbf{r}_{1\perp}-\mathbf{r}'_\perp). \tag{14}$$

Here $q(z,\mathbf{r}_\perp)$ is a nonrandom function of the coordinates, $N(\mathbf{r},\mathbf{r}')$ is the field line distribution function which satisfies the equation

$$\frac{\partial N}{\partial z'}+\frac{\partial}{\partial\mathbf{r}'_\perp}\cdot(\mathbf{A}(z',\mathbf{r}'_\perp)N)=0, \quad N\big|_{z=z'}=\delta(\mathbf{r}_\perp-\mathbf{r}'_\perp). \tag{15}$$

Expression (14) corresponds to the fact that for $D_\parallel\gg D_\perp$, the particle moves along the magnetic field line for a long time. The dependence of the function q on $\mathbf{r}_\perp$ describes the drift of the particle from the field line (an equation will be derived below for this function).

When substituting the trial function (14) into the first inequality (9), we use the obvious inequality $(\mathbf{b}\cdot\mathbf{g})^2\leqslant g_z^2+(\mathbf{g}\cdot\mathbf{A})^2$. The new functional obtained by making this substitution will be denoted by the letter L. In order to calculate this functional, we need to be able to calculate the averages of the product of the components of the vector **A** and the functions N, for which we can use perturbation theory if the following condition is satisfied:

$$AL_\parallel/L_\perp\ll 1. \tag{16}$$

Here A is the amplitude of the random field and $L_\parallel$ and $L_\perp$ are the correlation dimensions of the random field parallel and perpendicular to the average field. Retaining terms through the second order in A in the expression for L, we obtain

$$L=(D_{\parallel}-D_{\perp})A_{ij}(0,0)G_iG_j+D_{\perp}G^2+4(D_{\parallel}-D_{\perp})$$
$$\times\int d^3r'd^2r_{1\perp}G_iG_j\delta(z-z')B_{ij}(\mathbf{r}_{\perp}-\mathbf{r}_{1\perp})$$
$$\times\langle N(\mathbf{r},\mathbf{r}')\rangle\frac{\partial q(z-z',\mathbf{r}'_{\perp}-\mathbf{r}_{1\perp})}{\partial z}+2$$
$$\times\int dz'd^2r'_{\perp}d^2r''_{\perp}d^2r_{1\perp}d^2r_{2\perp}B_{ij}(\mathbf{r}_{1\perp}-\mathbf{r}_{2\perp})$$
$$\times G_iG_j\Bigg[D_{\parallel}\langle N(z,z',\mathbf{r}_{\perp},\mathbf{r}'_{\perp})N(z,z',\mathbf{r}_{\perp},\mathbf{r}''_{\perp})\rangle$$
$$\times\frac{\partial q(z-z',\mathbf{r}'_{\perp}-\mathbf{r}_{1\perp})}{\partial z}\frac{\partial q(z-z',\mathbf{r}''_{\perp}-\mathbf{r}_{2\perp})}{\partial z}$$
$$+D_{\perp}\left\langle\frac{\partial N(z,z',\mathbf{r}_{\perp},\mathbf{r}'_{\perp})}{\partial\mathbf{r}_{\perp}}\frac{\partial N(z,z',\mathbf{r}_{\perp},\mathbf{r}''_{\perp})}{\partial\mathbf{r}_{\perp}}\right\rangle$$
$$\times q(z-z',\mathbf{r}'_{\perp}-\mathbf{r}_{1\perp})q(z-z',\mathbf{r}''_{\perp}-\mathbf{r}_{2\perp})\Bigg]. \quad (17)$$

Here

$$B_{ij}(\mathbf{r}_{\perp})=\frac{1}{2}\int dz\ A_{ij}(z,\mathbf{r}_{\perp})$$

is the correlation function of the statistically uniform random magnetic field $A_{ij}(z,\mathbf{r}_{\perp})=\langle A_i(z,\mathbf{r}_{\perp})A_j(0,0)\rangle$ integrated over z. In the derivation of expression (17) we assumed that the characteristic scale of variation of the functions N and q with respect to z is much greater than the correlation dimension $L_{\parallel}$, which allowed us to perform a single integration over z'' in the second integral. Note that expression (17) is accurate if the field $A_k(z,\mathbf{r}_{\perp})$ is delta-correlated with respect to z and has Gaussian statistics.

In order to calculate the averages in the second integral in expression (17), we analyze the function $F(z,z',\mathbf{r}_{\perp},\mathbf{r}'_{\perp},\mathbf{r}''_{\perp})=N(z,z',\mathbf{r}_{\perp},\mathbf{r}'_{\perp})N(z,z',\mathbf{r}_{\perp},\mathbf{r}''_{\perp})$ for which an equation is derived from Eq. (15):

$$\frac{\partial F}{\partial z'}+\frac{\partial}{\partial\mathbf{r}'_{\perp}}\cdot(\mathbf{A}(z',\mathbf{r}'_{\perp})F)+\frac{\partial}{\partial\mathbf{r}''_{\perp}}\cdot(\mathbf{A}(z',\mathbf{r}''_{\perp})F)=0. \quad (18)$$

The first and second averages in the second integral in expression (17) have the following initial conditions:

$$F(z,z',\mathbf{r},\mathbf{r}'_{\perp},\mathbf{r}''_{\perp})\big|_{z=z'}=\delta(\mathbf{r}_{\perp}-\mathbf{r}'_{\perp})\,\delta(\mathbf{r}_{\perp}-\mathbf{r}''_{\perp}),$$
$$F(z,z',\mathbf{r},\mathbf{r}'_{\perp},\mathbf{r}''_{\perp})\big|_{z=z'}=\frac{\partial\delta(\mathbf{r}_{\perp}-\mathbf{r}'_{\perp})}{\partial\mathbf{r}_{\perp}}\cdot\frac{\partial\delta(\mathbf{r}_{\perp}-\mathbf{r}''_{\perp})}{\partial\mathbf{r}_{\perp}}. \quad (19)$$

An equation for the averaged function $F_0=\langle F\rangle$ may be obtained by using perturbation theory (for further details see Ref. 9):

$$\mathrm{sign}(z'-z)\frac{\partial F_0}{\partial z'}=B_{ij}(0)\left(\frac{\partial^2F_0}{\partial r'_{\perp i}\partial r'_{\perp j}}+\frac{\partial^2F_0}{\partial r''_{\perp i}\partial r''_{\perp j}}\right)$$
$$+2\frac{\partial^2}{\partial r'_{\perp i}\partial r''_{\perp j}}B_{ij}(\mathbf{r}'_{\perp}-\mathbf{r}''_{\perp})F_0. \quad (20)$$

This equation is accurate if the field $A_k(z,\mathbf{r}_{\perp})$ is delta-correlated with respect to z and has Gaussian statistics. Its solutions with the initial conditions (19) may be obtained by using a Fourier transformation with respect to $\mathbf{r}'_{\perp}$ and $\mathbf{r}''_{\perp}$:

$$\mathrm{sign}(z'-z)\frac{\partial F_{\mathbf{k}'_{\perp},\mathbf{k}''_{\perp}}}{\partial z'}=-\int\frac{d^2k_{\perp}}{(2\pi)^2}H_{ij}(\mathbf{k}_{\perp})$$
$$\times[(k'_{\perp i}k'_{\perp j}+k''_{\perp i}k''_{\perp j})F_{\mathbf{k}'_{\perp},\mathbf{k}''_{\perp}}$$
$$+2k'_{\perp i}k''_{\perp j}F_{\mathbf{k}'_{\perp}-\mathbf{k}_{\perp},\mathbf{k}''_{\perp}+\mathbf{k}_{\perp}}], \quad (21)$$

where the tensor $H_{ij}(\mathbf{k}_{\perp})$ is related to the correlation function of the random magnetic field by the expression

$$H_{mn}(\mathbf{k}_{\perp})=H(k_{\perp})\left(\delta_{mn}-\frac{k_{\perp m}k_{\perp n}}{k_{\perp}^2}\right)$$
$$=\int d^2r_{\perp}B_{mn}(\mathbf{r}_{\perp})\exp(-i\mathbf{k}_{\perp}\cdot\mathbf{r}_{\perp}),\quad m,n=1,2. \quad (22)$$

The initial conditions (19) in the Fourier representation have the form

$$F_{\mathbf{k}'_{\perp},\mathbf{k}''_{\perp}}\big|_{z=z'}=\exp[-i(\mathbf{k}'_{\perp}+\mathbf{k}''_{\perp})\cdot\mathbf{r}_{\perp}],$$
$$F_{\mathbf{k}'_{\perp},\mathbf{k}''_{\perp}}\big|_{z=z'}=-\mathbf{k}'_{\perp}\cdot\mathbf{k}''_{\perp}\exp[-i(\mathbf{k}'_{\perp}+\mathbf{k}''_{\perp})\cdot\mathbf{r}_{\perp}]. \quad (23)$$

For a random field statistically isotropic in the plane perpendicular to the average field, the solutions have the following form:

$$F_{\mathbf{k}'_{\perp},\mathbf{k}''_{\perp}}=\exp[-i(\mathbf{k}'_{\perp}+\mathbf{k}''_{\perp})\cdot\mathbf{r}_{\perp}-D_m(\mathbf{k}'_{\perp}+\mathbf{k}''_{\perp})^2|z-z'|],$$
$$F_{\mathbf{k}'_{\perp},\mathbf{k}''_{\perp}}=-\mathbf{k}'_{\perp}\cdot\mathbf{k}''_{\perp}\exp\{-i(\mathbf{k}'_{\perp}+\mathbf{k}''_{\perp})\cdot\mathbf{r}_{\perp}$$
$$-[D_m(\mathbf{k}'_{\perp}+\mathbf{k}''_{\perp})^2-2\gamma]|z-z'|\}, \quad (24)$$

or in the coordinate representation

$$F_0(z,z',\mathbf{r},\mathbf{r}'_{\perp},\mathbf{r}''_{\perp})=\langle N(z,z',\mathbf{r}_{\perp},\mathbf{r}'_{\perp})\rangle\,\delta(\mathbf{r}'_{\perp}-\mathbf{r}''_{\perp}),$$
$$F_0(z,z',\mathbf{r},\mathbf{r}'_{\perp},\mathbf{r}''_{\perp})=\exp(2\gamma|z-z'|)\frac{\partial}{\partial\mathbf{r}'_{\perp}}\cdot\frac{\partial}{\partial\mathbf{r}''_{\perp}}$$
$$\times\langle N(z,z',\mathbf{r}_{\perp},\mathbf{r}'_{\perp})\rangle\,\delta(\mathbf{r}'_{\perp}-\mathbf{r}''_{\perp}),$$
$$\langle N(z,z',\mathbf{r}_{\perp},\mathbf{r}'_{\perp})\rangle=\frac{1}{4\pi D_m|z-z'|}\exp\left[-\frac{(\mathbf{r}_{\perp}-\mathbf{r}'_{\perp})^2}{4D_m|z-z'|}\right]. \quad (25)$$

Here D_m and γ are the diffusion coefficient and the instability growth rate of the field lines, respectively, which are expressed in terms of the spectrum of the random magnetic field:

$$D_m=\frac{1}{2}\int\frac{d^2k_{\perp}}{(2\pi)^2}H(k_{\perp}),\quad \gamma=\frac{1}{2}\int\frac{d^2k_{\perp}}{(2\pi)^2}k_{\perp}^2H(k_{\perp}). \quad (26)$$

These coefficients describe the behavior of the magnetic field lines. The mean square of the displacement of the field line in the transverse direction is given by

$$\langle \mathbf{r}_\perp^2 \rangle = 4D_m |z-z'|, \tag{27}$$

and for the mean square of the distance between the nearest field lines in the transverse direction, we have

$$\langle (\mathbf{r}_{1\perp}-\mathbf{r}_{2\perp})^2 \rangle = \langle (\mathbf{r}_{1\perp}-\mathbf{r}_{2\perp})^2 \rangle|_{z=z'} \exp(2\gamma|z-z'|),$$
$$\langle (\mathbf{r}_{1\perp}-\mathbf{r}_{2\perp})^2 \rangle \ll L_\perp^2 . \tag{28}$$

Substituting the expressions (25) into Eq. (17), performing a single integration with respect to $\mathbf{r}'_\perp$, and converting to the $\mathbf{k}_\perp$ representation, we obtain

$$L = (D_\parallel - D_\perp) A_{ij}(0,0) G_i G_j + D_\perp G^2 + \int dz' d^2k_\perp G_i G_j H_{ij}(\mathbf{k}_\perp) \times \left[4(D_\parallel - D_\perp)\delta(z') \frac{\partial Q(z',\mathbf{k}_\perp)}{\partial z'} + 2D_\parallel \left(\frac{\partial Q(z',\mathbf{k}_\perp)}{\partial z'} \right)^2 + 2D_\perp k_\perp^2 \times \exp(2\gamma|z'|) Q^2(z',\mathbf{k}_\perp) \right]. \tag{29}$$

Here the function

$$Q(z,\mathbf{k}_\perp) = \int d^2r_\perp q(z,\mathbf{r}_\perp) \exp(-i\mathbf{k}_\perp \cdot \mathbf{r}_\perp)$$

is the Fourier transform of the function q. By varying this functional with respect to Q, we obtain the following expression:

$$D_\parallel \frac{\partial^2 Q}{\partial z^2} - k_\perp^2 D_\perp Q \exp(2\gamma|z|) = -(D_\parallel - D_\perp)\delta'(z). \tag{30}$$

Using the substitution $\zeta = \exp(\gamma|z|)$, we can reduce the equation to a Bessel equation of imaginary argument:

$$Q = -\frac{\operatorname{sign} z}{2} \left(1 - \frac{D_\perp}{D_\parallel} \right) K_0 \left(\frac{k_\perp}{\gamma} \sqrt{\frac{D_\perp}{D_\parallel}} \exp(\gamma|z|) \right) \Big/ K_0 \left(\frac{k_\perp}{\gamma} \sqrt{\frac{D_\perp}{D_\parallel}} \right). \tag{31}$$

Substituting this function into expression (29) and assuming that at the extremum the value of the integral is half that given by the first term in brackets, we obtain an upper bound on the effective transverse diffusion coefficient:

$$D_\perp^* \leqslant D_\perp - \frac{1}{2} \int \frac{d^2k_\perp}{(2\pi)^2} k_\perp H(k_\perp) \sqrt{D_\parallel D_\perp} K_0' \times \left(\frac{k_\perp}{\gamma} \sqrt{\frac{D_\perp}{D_\parallel}} \right) \Big/ K_0 \left(\frac{k_\perp}{\gamma} \sqrt{\frac{D_\perp}{D_\parallel}} \right). \tag{32}$$

Here the prime denotes differentiation with respect to the argument. The order-of-magnitude instability growth rate of the field lines is $\gamma \approx (1/2)\langle A^2 \rangle L_\parallel / L_\perp^2$ (we assume that the random field is concentrated in the main scale, which is of the order of the correlation dimension, so that the integrals (26) are determined by the amplitude of the random field on this scale) and thus in the most interesting case $\langle A^2 \rangle^2 L_\parallel^2 D_\parallel / L_\perp^2 D_\perp \gg 1$, the argument of the function K_0 is small in that part of the integration region where the expression in the integrand differs appreciably from zero. Using the asymptotic form of this function for $x \to 0$, $K_0(x) \approx -\ln x$, we obtain

$$D_\perp^* \leqslant D_\perp + \frac{1}{2} D_\parallel \gamma \int \frac{d^2k_\perp}{(2\pi)^2} H(k_\perp) \Big/ \ln \left(\frac{\gamma}{k_\perp} \sqrt{\frac{D_\parallel}{D_\perp}} \right). \tag{33}$$

5. CONCLUSIONS

In many cases, the variational approach can be used to estimate the limits of variation of the effective diffusion coefficient. If information is available on the spectrum of the random magnetic field and $D_\parallel > D_\perp$ holds, it follows from inequality (11) that the effective parallel diffusion coefficient lies within the range

$$D_\parallel / (1 + \langle A^2 \rangle) \leqslant D_\parallel^* \leqslant D_\parallel . \tag{34}$$

In the two-dimensional case, inequality (13) gives the following range of variation of the effective perpendicular diffusion coefficient:

$$D_\perp \leqslant D_\perp^* \leqslant D_\perp (1 + \langle A^2 \rangle). \tag{35}$$

Thus, in the limit $\langle A^2 \rangle \ll 1$ in these cases the effective diffusion coefficients will be close to the corresponding local coefficients. In the three-dimensional case, the inequality (33) gives an upper estimate of the effective diffusion coefficient:

$$D_\perp^* \leqslant D_\perp + D_\parallel D_m \gamma \Big/ \ln \left(\gamma L_\perp \sqrt{\frac{D_\parallel}{D_\perp}} \right) \approx D_\perp + \frac{D_\parallel \langle A^2 \rangle^2 L_\parallel^2}{4L_\perp^2} \Big/ \ln \left(\frac{\langle A^2 \rangle L_\parallel}{L_\perp} \sqrt{\frac{D_\parallel}{D_\perp}} \right). \tag{36}$$

Here the equality is only approximate because of the assumptions made to derive the inequality (36). The right-hand side of inequality (36) is of the same order of magnitude as the estimated transverse diffusion coefficient obtained in Refs. 10 and 11. Let us assume that τ is the time when the particle ''forgets'' the field line. As it moves along the field line as a result of diffusion, the particle covers the distance $\sqrt{D_\parallel \tau}$ in this time and the transverse displacement as a result of ''wandering'' of the field line will be $D_m \sqrt{D_\parallel \tau}$. Thus, in order of magnitude the transverse diffusion coefficient is given by

$$D_\perp^* \approx D_m \sqrt{D_\parallel / \tau}. \tag{37}$$

Over the time τ the particle is displaced by the distance $\sqrt{D_\perp \tau}$ in the transverse direction as a result of diffusion. This distance increases as a result of the instability of adjacent field lines (see formula (28)). Equating this to the transverse correlation scale, we obtain

$$D_\perp \tau \exp(2\gamma\sqrt{D_\parallel \tau}) \approx L_\perp^2 . \quad (38)$$

Solving this equation approximately for τ, we have

$$\tau \approx \frac{1}{D_\parallel \gamma^2} \ln^2\left(\gamma L_\perp \sqrt{\frac{D_\parallel}{D_\perp}} \right). \quad (39)$$

Substituting this expression into formula (37), we obtain an expression for the transverse diffusion coefficient which agrees with the right-hand side of inequality (36). Note that for $D_\perp \to 0$ this value is lower than the transverse diffusion coefficient calculated using various approximate methods.[12–14].

This work was supported by the ''Astronomy'' Program (Grant No. 2-149).

*)E-mail:vp@cosray.izmiran.troitsk.su

[1] J. R. Jokipii, Rev. Geophys. Space Phys. **8**, 27 (1971).
[2] J. J. Quenby, Space Sci. Rev. **37**, 201 (1984).
[3] I. N. Toptygin, *Cosmic Rays in Interplanetary Magnetic Fields*, Boston; Kluwer Academic, Dordrecht (1985) [Russian original, Nauka, Moscow (1983)].
[4] V. S. Berezinskiĭ, S. V. Bulanov, V. L. Ginzburg, V. A. Dogel', and V. S. Ptuskin, *Astrophysics of Cosmic Rays*, edited by V. L. Ginzburg, North–Holland, Amsterdam (1990) [Russian original, Nauka, Moscow (1990)].
[5] A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 641 (1970) [Sov. Phys. JETP **32**, 348 (1970)].
[6] A. M. Dykhne, Zh. Éksp. Teor. Fiz. **52**, 264 (1967) [Sov. Phys. JETP **25**, 170 (1967)].
[7] M. I. Shvidler, *Statistical Hydrodynamics of Porous Media* [in Russian], Nedra, Moscow (1981).
[8] S. M. Kozlov, Usp. Fiz. Nauk **40**, 61 (1985) [*sic*].
[9] J. R. Jokipii, Astrophys. J. **183**, 1029 (1973).
[10] A. B. Rechester and M. N. Rosenbluth, Phys. Rev. Lett. **40**, 38 (1978).
[11] E. G. Klepach, V. S. Ptuskin, and L. G. Chuvil'gin, Izv. Ross. Akad. Nauk, Ser. Fiz. **57**(7), 86 (1993).
[12] B. B. Kadomtsev and O. P. Pogutse, in *Proceedings of the Seventh International Conference on Plasma Physics and Controlled Nuclear Fusion Research*, Vienna, 1978, Vol. 1, p. 649.
[13] A. A. Galeev and I. M. Zeleny, Physica (Amsterdam) **20**, 90 (1981).
[14] L. G Chuvilgin and V. S. Ptuskin, Astron. Astrophys. **279**, 278 (1993).

Translated by R. M. Durham

New classes of exact solutions in inflationary cosmology

V. M. Zhuravlev, S. V. Chervon,*) and V. K. Shchigolev

Ulyanovsk State University, 432700 Ulyanovsk, Russia
(Submitted 21 May 1997)
Zh. Éksp. Teor. Fiz. **114**, 406–417 (August 1998)

The problem of determining a representation of the self-interaction potential in the form of a time dependence of the field potential energy which admits the existence of an inflationary regime and the transition of evolution to a Friedmann regime of asymptotic expansion is investigated within a cosmological model with a self-interacting scalar field. A variational formulation of the slow-roll concept is introduced, and, on the basis thereof, an exact solution is constructed for the evolution of the scale factor and the form of the self-interaction potential. A method based on representing the Einstein equations in the form of a linear second-order equation is developed for constructing and analyzing exact cosmological solutions of these equations. Selected types of potentials and the corresponding evolutions of the universe are investigated. © *1998 American Institute of Physics.* [S1063-7761(98)00308-4]

1. INTRODUCTION

In this paper we investigate general properties of the evolution of a homogeneous isotropic universe on the basis of an analysis of the system of Einstein equations for a self-interacting scalar field. One of the important requirements imposed on early-universe theories by present-day observational data is the existence of an inflationary phase in the evolution of the universe with an exponentially increasing scale factor.[1] However, the physical causes responsible for the inflationary phase and its subsequent termination within a certain stage of the evolution of the universe have not been fully elucidated. There are several different hypotheses in this regard. It is therefore important to ascertain the general properties of the self-interacting-scalar-field model which characterize the inflationary phase and distinguish this field from the background of the general evolution of the universe over a period beginning with the preinflationary epoch and ending with later epochs in the evolution of the universe. It is reasonable to hope that this information can be used to elucidate the actual causes of the onset and termination of inflation. The formulation of the problems and the methods used to investigate them in the present study draw heavily on the method for fine tuning of the potential in Ref. 2 and illustrate its heuristic capabilities.

The inflationary phase of the (isotropic and homogeneous) universe is customarily understood to be the period $t_i<t<t_f$ of its early evolution, beginning at the time t_i and ending at the time t_f, when the scale factor $K(t)$ increased at an accelerated rate, i.e.,

$$\ddot{K}=K(H^2+\dot{H})>0$$

(see Ref. 3). In this setting the term subinflation is used when $\dot{H}<0$, standard or exponential inflation when $\dot{H}=0$, and superinflation in the case of $\dot{H}>0$.

We take a special look at the cases in which the law governing the variation of the scale factor $K(t)$ has one of the forms

$$K(t)=K_0t^m, \quad m>1, K_0>0, \tag{1}$$

$$K(t)=K_0e^{H_0t}, \quad H_0>0, \tag{2}$$

$$K(t)=K_0(t_*-t)^n, \quad n<0, \quad t_*>t_i, \tag{3}$$

which ensure fulfillment of the general inflation condition $\ddot{K}>0$. The scale factor (1) corresponds to subinflation and is called power-law inflation. The case of superinflation (3) is unattainable within the scope of the theory of a self-interacting scalar field.

The existence of an inflationary period of expansion of the universe is necessary for solving the horizon (homogeneity), flatness, and relic-monopole problems, which have been discussed in detail, for example, in the original work of Guth[4] (see also Linde's book[1]). The specific mechanism driving inflation in the theory with a self-interacting scalar field (inflaton field) is usually identified with the form of the functional dependence of the effective self-interaction potential of the scalar field on the field itself: $V=V(\phi)$. An important factor in substantiating the existence of such a mechanism is the slow-roll regime, i.e., a regime in which the variation of the field ϕ [and the potential $\mathbf{V}(\phi)$] was relatively slow, so that the ''kinetic'' energy of the scalar field can be disregarded: $\dot{\phi}^2 \ll V(\phi)$ (Ref. 5).

In the standard model (2) the self-interaction potential is specified as a quadratic function of ϕ:

$$V(\phi)=\mu^2\phi^2+b.$$

In this case the field varies linearly with time: $\phi \propto t$, whereas the scale factor increases exponentially: $K(t)\propto e^{Ht}$ (Refs. 1 and 6).

A new approach to the construction of exact solutions in inflationary cosmology (without any reliance on the slow-roll regime) has been developed in Refs. 2, 7, and 10. The method can be broadly summarized as follows. We specify the regime of evolution of the scale factor $K=K(t)$ and determine the evolution of the self-interaction potential $V=V(t)$ in such a way as to ensure the regime assigned to

$K=K(t)$. We then determine the rate of change of the scalar field $\dot{\phi}(t)$. Next, integrating over t, we find the law of evolution of the scalar field $\phi=\phi(t)$, which is also matched to the regime chosen for the scale factor. We ultimately obtain a parametric expression for $V=V(\phi)$, which then represents the self-interaction potential ''fine-tuned'' to the exact solution assigned.

In particular, in Ref. 2 the self-interaction potentials were plotted as a function of the scale factor $K(t)$ corresponding to the inflationary regimes (1) and (2). Also, in Ref. 2, $V(\phi)$ was obtained in a general parametric form for certain types of large-t asymptotic behavior of $K(t)$ which ensure regime (1) or (2).

A somewhat different approach to the construction of exact solutions in cosmological inflation models was presented by Barrow.[11] In his approach[11] the evolution of the scalar field $\phi=\phi(t)$ is first specified, and then the evolution of the scale factor $K=K(t)$ and the potential, which depends explicitly on the scalar field ϕ, are determined.

We also mention the work of Maartens *et al.*,[12] who used a method similar to fine tuning of the potential, but with a special parameter introduced in place of time. This approach might well be practical for the solution of some problems, but it complicates the process of obtaining and analyzing solutions.

Here we follow the general scheme of the method for fine tuning of the potential and submit two new techniques for constructing classes of exact inflationary solutions for a homogeneous isotropic universe. One technique entails a variational formulation of the conditions underlying the slow-roll regime, and the other rests on the possibility of specifying the self-interaction potential as a function of time. We investigate the general implications of such an approach and present new exact solutions.

2. EXACT SOLUTION AND SELF-INTERACTION POTENTIAL FOR THE SLOW-ROLL REGIME

The Einstein equations for a homogeneous isotropic universe and an arbitrary form of the self-interaction potential $V(\phi)$ of the scalar field $\phi(t)$ can be represented in the class of Friedmann metrics by two equations:[2]

$$V(t)=\frac{1}{\kappa}\left(\Lambda+\frac{\ddot{K}}{K}+2\frac{\dot{K}^2}{K^2}+\frac{2\epsilon}{K^2}\right), \tag{4}$$

$$\phi(t)=\pm\sqrt{\frac{2}{\kappa}}\int\sqrt{-\frac{d^2\ln K}{dt^2}+\frac{2\epsilon}{K^2}}\,dt+\phi_0, \tag{5}$$

where κ is the gravitational constant, Λ is the cosmological constant, and ϕ_0 is an integration constant. This system of equations provides the basis of the method for fine tuning of the potential. As mentioned in the Introduction, once the evolution of the scale factor has been specified, this method can be used to find the appropriate form of the self-interaction potential to achieve a given regime of evolution of the universe.

The method has been used previously[2,8] to analyze inflationary regimes and to demonstrate the existence of a large number of self-interaction potentials admitting such a regime of evolution. The enormous diversity of potentials leading to inflation compels us to look for an additional principle that might be used to isolate the potential that actually occurred in the early stages of evolution of the universe. Considering the special importance of the existence of the slow-roll regime for an inflationary scenario, we look into a nonstandard formulation of the slow-roll regime. We define the slow-roll regime with the aid of the variational principle of minimum variation of the scalar field ϕ with variation of the scale factor $K(t)$ and, as a consequence, minimum variation of the values of the potential $V(\phi)$. On the basis of Eq. (5) this condition can be specified by the variational equation

$$\delta\phi=\int_{t_1}^{t_2}\delta\sqrt{-\frac{d^2\ln K}{dt^2}+\frac{2\epsilon}{K^2}}\,dt=0, \tag{6}$$

where $t_i\leqslant t_1<t_2\leqslant t_f$. This condition literally means that the evolution of the scale factor must be such that the difference in the values of the field ϕ in any finite time interval $[t_1,t_2]$ would be the smallest among all other possible evolutions. The variation of the value of $V(\phi)$ is also the smallest in this case.

If we introduce the notation

$$F(t)=\left(-\frac{d^2\ln K}{dt^2}+\frac{2\epsilon}{K^2}\right)^{-1/2},$$

the Euler–Lagrange equations corresponding to the variational problem (6) assume the form of a pair of equations for $F(t)$ and $K(t)$:

$$-\frac{d^2\ln K}{dt^2}+\frac{2\epsilon}{K^2}=\frac{1}{F^2}, \tag{7}$$

$$\frac{d^2F}{dt^2}+\frac{2\epsilon}{K^2}F=0. \tag{8}$$

It is simplest to find exact solutions of this system for the case $\epsilon=0$, which corresponds to a spatially flat Friedmann universe. In this case, from Eqs. (7) and (8) we have:

$$-\frac{d^2\ln K}{dt^2}=\frac{1}{F^2},\quad \frac{d^2F}{dt^2}=0.$$

The solution of these equations is given by the functions

$$K(t)=k_0(a_0t+b_0)^{1/a_0^2}e^{c_0t},\quad F(t)=a_0t+b_0, \tag{9}$$

where a_0, b_0, c_0, and k_0 are arbitrary constants. Substituting the solution (9) into Eqs. (4) and (5), we obtain the following results for the field and the potential:

$$\phi(t)=\sqrt{\frac{2}{\kappa}}\frac{1}{a_0}\ln(a_0t+b_0)+\phi_0, \tag{10}$$

$$V(\phi,c_0)=\frac{1}{\kappa}\left[\Lambda+3c_0^2+\left(\frac{3}{a_0^2}-1\right)e^{-2\alpha(\phi-\phi_0)}+\frac{6c_0}{a_0}e^{-\alpha(\phi-\phi_0)}\right], \tag{11}$$

where $\alpha=a_0\sqrt{\kappa/2}$. It is interesting to note that the solution (9)–(11) generalizes a solution obtained previously[13] by generating new solutions using invariant transformations of ϕ,

K, and V, which do not alter the equations of the standard inflation model. As should be expected, this solution describes an inflationary regime of the exponential or power-law type or a combination thereof.

Parallels between the solution (10), (11) and the equation of state of matter can be drawn on the basis of the analogy between the theories of a self-interacting-scalar-field and an ideal fluid. This analogy is expressed, in particular, in the fact that by comparing the energy-momentum tensors of these two models one can formally calculate the pressure p and the energy density ρ of an ideal fluid in terms of the parameters of the scalar-field model according to the rule

$$T_4^4=\rho=\frac{1}{2}\,\dot{\phi}^2+V(\phi), \tag{12}$$

$$-T_1^1=-T_2^2=-T_3^3=p=\frac{1}{2}\,\dot{\phi}^2-V(\phi). \tag{13}$$

Substituting the expressions for the self-interaction potential $V(\phi)$ from Eq. (11) and the field ϕ from Eq. (10), we obtain the relations

$$\rho=\frac{1}{\kappa}\left[\Lambda+3c_0^2+\frac{3}{a_0^2}e^{-2\alpha(\phi-\phi_0)}+\frac{6c_0}{a_0}e^{-\alpha(\phi-\phi_0)}\right],$$

$$p=-\frac{1}{\kappa}\left[\Lambda+3c_0^2+\left(\frac{3}{a_0^2}-2\right)e^{-2\alpha(\phi-\phi_0)}+\frac{6c_0}{a_0}e^{-\alpha(\phi-\phi_0)}\right].$$

Eliminating the field ϕ from these equations, we obtain an effective equation of state of matter in the form

$$p=-\rho+\frac{2a_0^2}{\kappa}\left(-c_0+\sqrt{\frac{1}{3}(\kappa\rho-\Lambda)}\right)^2. \tag{14}$$

When the class of solutions (9)–(11) is extrapolated to small and large times, i.e., beyond the limits of the inflationary phase $t_i\leqslant t_1<t_2\leqslant t_f$, the following characteristics are readily established. The solution under consideration always begins from a singular state, i.e., for any model parameters (Λ,a_0,c_0,b_0) there is a time $t_0=-b_0/a_0$ in the history of the evolution of the scale factor when $K(t_0)=0$, after which the scale factor increases at an accelerated rate for some time. A natural transition to Friedmann expansion does not take place at larger times, and this is a drawback of the solutions under analysis. The transition to the Friedmann regime requires that $1/a_0^2=2/3$ and $c_0=0$. It is therefore obvious that in the absence of the cosmological constant Λ in Eq. (4) its role is assumed by c_0, and the condition $c_0=0$ corresponds to the standard departure from the inflationary regime.[1]

Again we emphasize that the analytical solution (9)–(11) corresponds to a potential tuned exactly to the slow-roll regime in the variational formulation.

It is a well-known fact that the pioneering work on inflation (see the reviews in Refs. 1, 5, and 6) treated a truncated system of Einstein and scalar-field equations, with the slow-roll regime defined so as to eliminate the second time derivative of the scalar field $\ddot{\phi}$ from the equations, a device justified by the condition $\dot{\phi}^2\ll V(\phi)$. There are no significant constraints on the self-interaction potential in this approach. In our case we have obtained an exact representation of the potential in the explicit form (11). We note once again that a solution analogous to (11) has been obtained previously,[13] but without any mention of the slow-roll regime *per se*. It is easily verified that the condition $\dot{\phi}^2\ll V(\phi)$ is satisfied in a certain time period $t_i\leqslant\Delta t<t_f$ for the exact solution (9)–(11).

3. GENERATION OF EXACT SOLUTIONS FOR A GIVEN POTENTIAL $V=V(t)$

The problem of analyzing the interrelationship between the scenario of evolution of the universe and the form of the self-interaction potential can be treated in time scales greater than the inflationary phase. In this section, with a view toward completeness, we discuss the overall evolution of the universe.

In the standard approach the self-interaction potential of the scalar field appears in the Einstein equations as the function $V=V(\phi)$. In our approach $V(\phi)$ is actually replaced by $V(t)$, which must be interpreted as an effective potential of material fields. We note that in the present case of a homogeneous and isotropic universe the representation of the self-interaction potential by $V=V(t)$ does not conflict with the general variational problem of the derivation of the Einstein equations, where significant use is made of the functional dependence $V(\phi)$, since each of the functions $V(t)$ and $\phi(t)$ corresponds uniquely to a particular function $V(\phi)$. We call the function $V(t)$ the evolution of the potential energy or the history of the potential, underscoring the departure of our approach from the standard approach, in which the dependence $V=V(\phi)$ is fixed.

Let us consider the problem of finding all possible types of evolution of the universe for a fixed history of the potential $V=V(t)$, which describes the time variation of the potential energy of the material fields. Stated in the form indicated, the problem has exact solutions, whose construction reduces to the integration of linear equations in the case of $\epsilon=0$. We write Eq. (4) as an equation for the function $Z(t)=K^3$:

$$-\ddot{Z}+3(\kappa V(t)-\Lambda)Z-6\epsilon Z^{1/3}=0. \tag{15}$$

In the case of a spatially flat universe ($\epsilon=0$) this equation acquires the form of the ordinary linear differential equation

$$-\ddot{Z}+3(\kappa V(t)-\Lambda)Z=0, \tag{16}$$

which has the same form as the Schrödinger equation for the motion of a quantum particle with a potential energy $U(t)=3\kappa V(t)$ and a self-energy $E=3\Lambda$ in one-dimensional space. Now the function $Z(t)$ assumes the role of the wave function of the particle. We wish to examine this case in more detail.

We assume that one of the solutions of this equation for a fixed dependence $V=V(t)$ has been found. We denote it by $Z_1(t)$, whereupon a second linearly independent solution of this equation $Z_2(t)$ for a fixed $U(t)$ is easily found, because any two linearly independent solutions of Eq. (16) are related by the equation

$$Z_1\dot{Z}_2 - Z_2\dot{Z}_1 = W_0, \tag{17}$$

where W_0 is a constant. From this result we obtain

$$Z_2 = Z_1\left(Q_0 + W_0\int \frac{dt}{Z_1^2}\right), \tag{18}$$

where Q_0 is an integration constant. All solutions of Eq. (16) corresponding to an assigned fixed potential $U(t)$ can be found by varying the constants Q_0 and W_0.

For example, the solution (9) for $K(t)$ in the problem with a minimally varying field leads to the solutions

$$K_2(t) = K_0(t)\left[Q_0 + W_0\int K^{-6}(t)dt\right]^{1/3}$$
$$= k_0(a_0t+b_0)^{1/a_0^2}e^{c_0t}\left[Q_0 + k_0^6W_0 \times \int dt(a_0t+b_0)^{-6/a_0^2}e^{-6c_0t}\right]^{1/3}. \tag{19}$$

In the case of pure power-law inflation $c_0 = 0$ in the solution (9) with the minimally varying scalar field we obtain a general class of solutions for $V(t)$ fixed in (10):

$$K_2^{(p)}(t) = k_0(a_0t+b_0)^{1/a_0^2}\left[Q_0 + k_0^6W_0\int dt(a_0t+b_0)^{-6/a_0^2}\right]^{1/3}$$
$$= k_0(a_0t+b_0)^{1/a_0^2}[Q_0 + k_0^6W_0A(a_0t+b_0)^{-6/a_0^2+1}]^{1/3}, \tag{20}$$

where $A = a_0^2/(a_0^2-6)$.

Note that the solutions (19) and (20) are new exact solutions for the parametric dependence $V = V(\phi)$, i.e., $V = V(t)$ and $\phi = \phi(t)$, indicated in the solution (10)–(11).

The proposed method based on (17) carries over without too much difficulty to the case of $\epsilon \neq 0$. It is necessary here to fix the function

$$Q(t) = 3(\kappa V(t) - \Lambda) - 6\epsilon Z^{-2/3}$$

as a function of time. The potential $V(t)$ then depends on the form of the solution. The case of $\epsilon \neq 0$ requires a separate investigation.

The linearity of Eq. (16) for $Z(t)$ permits the formulation of a problem in eigenfunctions and eigenvalues, the role of which is taken here by the cosmological constant, provided this equation is supplemented with homogeneous initial conditions. For example, we can investigate all possible evolutions for a fixed $V(t)$ which begin at a certain time t_0 from the state $K(t) = Z(t) = 0$ and return to the same singular state at another time $t_1 > t_0$. Such scenarios of the evolution of the universe can be mapped to oscillating solutions:[15] the universe originated and disappeared in the time period (t_0, t_1). The corresponding problem appears as follows:

$$-\ddot{Z} + 3(\kappa V(t) - \Lambda)Z = 0, \tag{21}$$

$$Z|_{t=t_0} = 0, \quad Z|_{t=t_1} = 0. \tag{22}$$

This problem has the same form as quantum-mechanical problems for a discrete spectrum. If the initial conditions are set at times $t_0 = -\infty$ and $t_1 = +\infty$, and the potential energy is a smooth function of time, the universe can evolve from a state other than a singular state and pass again into a nonsingular state inasmuch as the energy densities and other physical characteristics of matter (field) are finite. Other types of homogeneous initial conditions are also possible. The fact that there is only a finite or denumerable set of admissible values of the cosmological constant for each problem of this kind can shed light on the issue of the true value of the cosmological constant and the physical reasons for it.

4. ANALYSIS OF THE EVOLUTION OF THE UNIVERSE FOR VARIOUS TYPES OF POTENTIALS

The determination of the features of the origin and evolution of various inflationary regimes can be pursued on the basis of the representation (16) in the example of a series of models that are simple from the standpoint of constructing solutions but are interesting from the standpoint of the behavior of the potentials. In the present study we have chosen the following model potentials:

$$3\kappa V(t) = 2t^2, \tag{a}$$

$$3\kappa V(t) = \frac{m}{t^2}, \quad m = \text{const}, \tag{b}$$

$$3\kappa V(t) = -\frac{2\lambda_0}{\cosh^2(\lambda_0 t)}. \tag{c}$$

We investigate the solutions for the potentials (a), (b), and (c) from the viewpoint of the possible existence of inflationary regimes and their transition to a Friedmann phase of evolution.

(a) It is readily verified that the potential (a) admits oscillating solutions corresponding to the statement of the problem in the form (21), i.e., laws of evolution of the scale factor such that the universe goes from a singularity in the limit $t \to -\infty$ to a new singularity in the limit $t \to +\infty$. Thus, the solutions of Eq. (16) with the zero boundary conditions for $t \to \pm\infty$ have the form

$$Z(t) = H_i(t)e^{-t^2/2},$$

where $H_i(t)$ are Hermite polynomials. For example, the simplest solution has the form $Z(t) = d_0^3 \exp\{-t^2/2\}$. In this case

$$K(t) = d_0 \exp\{-t^2/6\}, \quad \phi(t) = \pm\sqrt{\frac{2}{3\kappa}}\,t + \phi_0,$$

$$V(\phi) = (\phi(t) - \phi_0)^2.$$

The value of the cosmological constant ensuring this evolution regime is $\Lambda = 1/3$ (in appropriate units of measure for the problem). For this particular solution the condition

$$\ddot{K} = (d_0/3)\exp\{-t^2/6\}\{t^2/3 - 1\} > 0$$

implies the start of inflation at $t > \sqrt{3}$ and no exit from the inflationary regime.

Other Hermite polynomials correspond to higher absolute values of $\Lambda = 1, 5/3, \ldots$. The evolution regimes in this case are such that the universe passes through a singular state several times. It is readily verified that the inflationary phase

for these oscillating solutions begins at $t\to-\infty$, where the scale factor has a zero minimum, and ends at a time t_0 corresponding to the point of inflection of the function $K=K(t)$.

(b) Potentials of the type (b) for $m\geqslant -1/4$ are interesting because in the case of $\Lambda=0$ they describe all possible types of power-law evolution of the scale factor, including power-law inflation. In fact, the general solution of Eq. (21) for arbitrary $m>0$ and $\Lambda=0$ has the form

$$Z(t)=C_1t^{\alpha}+C_2t^{\beta}, \tag{23}$$

where α and β are noncoinciding solutions of the algebraic equation $x^2-x-m=0$, i.e.,

$$\alpha=\frac{1}{2}+\sqrt{\frac{1}{4}+m}>0, \quad \beta=\frac{1}{2}-\sqrt{\frac{1}{4}+m}<0.$$

Hence it is evident that in the case of $C_1>0$ and $C_2>0$ the solution (23) is positive at $t>0$ and has one minimum at a point $t_0>0$ and that after passage through this minimum, power-law inflation begins at a time $t_1>t_0$, asymptotically approaching the $t^{\alpha/3}$ regime. The asymptotic power-law inflation regime corresponds to the requirement $\alpha>3$, so that $m>6$. We note that at times prior to t_1 the field ϕ is imaginary and that the evolution regime physically achieved in the self-interacting-scalar-field model begins precisely at t_1. To avoid this problem, it must be required that the conditions $C_1>0$ and $C_2\leqslant 0$ hold. In this case evolution begins from a singular state at a time t_S and makes an immediate transition to power-law inflation.

Remarkably, the only regime of power-law evolution of the scale factor that does not lead to the potential (b) is Friedmann expansion with the ultimately rigid state of matter $p=\rho$, for which $K(t)\propto t^{1/3}$, so that $Z\propto t$ and $V(t)=0$ for $\Lambda=0$. The Friedmann expansion regime $K(t)\propto t^{1/3}$ sets in as $t\to\infty$ provided the potential tends to zero more rapidly than $1/t^2$ as it decreases. An example of this kind of behavior is afforded by one of the possible forms of $K=K(t)$ based on evolution according to the function

$$Z(t)=z_0+\frac{t^2}{1+t}.$$

At $t>-1$ the scale factor passes through a minimum value $z_0^{1/3}$ at the time $t=0$, after which inflation begins. Inflation ends when the point of inflection is reached, and then the evolution asymptotically settles into the Friedmann regime $K(t)\propto t^{1/3}$ as $t\to+\infty$. The self-interaction potential in this case is

$$V(t)=\frac{\Lambda}{\kappa}+\frac{2}{3\kappa(1+t)^2}\,\frac{1}{t^2+z_0t+z_0},$$

whence it follows that

$$V(t)\to\frac{2}{3\kappa}\,\frac{1}{t^4} \text{ as } t\to\infty.$$

For transition to the Friedmann regime with dustlike matter, i.e., when $p=0$, it is sufficient that $V(t)\to 2/t^2$ as $t\to\infty$. For any other equation of state of the type $p=\gamma\rho$, where $\gamma=\text{const}$, it is sufficient that $V(t)\to m/t^2$ for $m\geqslant -1/4$ as $t\to\infty$. The case of $m=-1/4$ corresponds to a radiation-dominated state of matter.

The potential (b) also leads to oscillating solutions for $m<0$, which correspond to $\Lambda<0$. For example, the general solution of Eq. (16) for the potential (b) in the case of $m=-2$ and $\Lambda<0$ has the form

$$Z(t)=A\left(k-\frac{1}{t}\right)e^{kt}+B\left(-k-\frac{1}{t}\right)e^{-kt}, \quad k=\sqrt{-3\Lambda}. \tag{24}$$

The set of solutions of Eq. (24) includes one that satisfies the condition $Z(\pm\infty)=0$. It is the solution corresponding to $\Lambda=0$ and has the form $Z=A/|t|$, $A>0$.

(c) To discern certain features of the limitations imposed on the growth rate by the condition that there be a transition to Friedmann expansion, we give additional consideration to the history of the potential (c). Like the potential (a), it increases as $t\to\infty$, but only to a constrained value equal to zero. In this case the solution for $Z(t)$ with $\Lambda<0$ has the form

$$Z(t)=A(\lambda-\lambda_0\tanh(\lambda_0t))e^{\lambda t}+B(\lambda+\lambda_0\tanh(\lambda_0t))e^{-\lambda t}.$$

Here $\lambda^2=-3\Lambda>0$. When $\lambda=\lambda_0$, this potential corresponds to oscillatory evolution of the form

$$Z(t)=\frac{C}{\cosh(\lambda_0t)}, \quad K(t)=C^{1/3}\cosh^{-1/3}(\lambda_0t),$$

where C is an arbitrary constant. This is a unique solution for the potential under consideration and $\lambda=\lambda_0$, which corresponds to a unique bound state. As in the case of the potential (a), it describes an inflationary regime in the interval $(-\infty,t_0)$, where t_0 is the point of inflection of the function $K(t)$.

In the case of $\lambda=\Lambda=0$ the solution is the function

$$K(t)=C\tanh^{1/3}\lambda_0t,$$

which describes the departure of the universe from a singular state at the time $t=0$ and its asymptotic transition to the stationary state $K=C=\text{const}$ as $t\to+\infty$.

In the case of $\lambda>\lambda_0$, $A>0$, and $B=0$ we have the solution

$$Z(t)=A(\lambda-\lambda_0\tanh(\lambda_0t))e^{\lambda t},$$

$$K(t)=A^{1/3}e^{\lambda t/3}(\lambda-\lambda_0\tanh(\lambda_0t))^{1/3},$$

which describes evolution without singularities and with the onset of the inflationary regime at a time $t_0>0$.

The solutions corresponding to $\Lambda>0$ are oscillating solutions that pass repeatedly through the value $Z=0$. It is evident that asymptotic Friedmann expansion does not occur for this potential.

Comparing the solutions obtained for the three types of potentials, we arrive at the following conclusions. First, the potentials that increase as $t\to\infty$ [potentials (a) and (c)] do not allow any transition to the Friedmann regime if the growth rate of the potential energy exceeds the rate at which the function mt^{-2} with $m=-1/4$ approaches zero. Transition to the Friedmann regime requires that the potential de-

crease with time according to a power law of the type mt^{-2} with $m>0$ or that it increase by an analogous law with $-1/4<m<0$. Second, all three types of potentials exhibit the existence of the inflationary regime, corroborating the conclusion that this regime is not selective with regard to the form of the potential. All this indicates that the most preferred models are those of the type (b), which require slight modifications to ensure transition to the required Friedmann regime. For example, one such modification is the introduction of a weak dependence of m in the equation for the potential (b) on the temperature T of matter

$$V(t,T)=\frac{m(T)}{t^2} \tag{25}$$

such that $m(T)>0$ near the minimum of $K(t)$ and $m(T)\to 2$ in the limit $t\to\infty$. For example, $m(T)$ can be an abruptly varying function that characterizes phase transitions in early universe matter. The potential (11) falls in this category in the case $c_0=0$. Consequently, these potentials (b) satisfy the slow-roll principle in the variational formulation set forth in Sec. 2.

It should also be noted that the simplicity of Eq. (25) for the potential is preserved only when it is written in the form $V=V(t)$. If the form of the function $V=V(\phi,T)$ is written on the basis of the solutions obtained for $K(t)$ and $\phi(t)$, this function acquires an intricate form and depends significantly on the integration constants of Eq. (16), which, in turn, are determined by the initial conditions for $K(t)$ and $\phi(t)$. The latter demonstrates the preferability of $V=V(t)$ over $V=V(\phi)$ for analyzing the dynamics of a model.

5. CONCLUSION

We have found new classes of exact solutions for a self-consistent system of gravitating scalar fields with self-interaction within the cosmology of a homogeneous isotropic universe. We have analyzed aspects of this model that have bearing on the determination of the physical conditions restricting the admissible form of the self-interaction potential and on the determination of the general form of evolution of the scale factor from the standpoint of the existence of an inflationary phase in the evolution of the universe and some of its other fundamental characteristics.

1. We have shown that the definition of the slow-roll regime, which is actively utilized in inflation, is amenable to a variational formulation. In this formulation we have obtained an exact solution for the standard model of inflation with a self-interacting scalar field and have given an equation of state of matter corresponding to this solution.

2. We have proposed a method for generating exact solutions with a fixed history of the potential $V=V(t)$. We have used this method to formulate an eigenfunction/eigenvalue problem for a flat Friedmann universe with the role of the eigenvalues taken by the cosmological constant. We have shown that in order to analyze the character of the evolution of the universe, it is simpler to analyze the dynamics of the model using a representation of the potential in the form of its history $V=V(t)$ than a representation in the form of a function of the field $V=V(\phi)$.

3. We have analyzed several characteristic types of histories of the potential and the corresponding evolutions of the scale factor. Our analysis shows that the formal analog of the Einstein equations in the form of Schrödinger equation (15) or (16) proposed in this paper can be used to analyze in detail the behavior of various physical factors in models of self-interacting scalar fields and to reveal a potential selection criterion on the basis of physical conceptions regarding the character of the evolution of the universe in large time scales with inflation (subinflation) as one of the stages, as well as the conditions for passage of the universe into the Friedmann regime. It follows from our analysis that the most realistic model of the history of the potential energy is a model of the form (25).

In closing, we gratefully acknowledge the reviewer's valuable comments, which have contributed to the improvement of the paper. One of us (Chervon) wishes to thank the members of the Gravitation Seminar at the Shternberg State Institute of Astronomy and A. A. Starobinskiĭ for a fruitful discussion of some questions that were addressed in the paper.

This work was carried out under the aegis of the State Scientific-Technical Program ''Astronomy: Fundamental Research of the Cosmos,'' Section ''Cosmomicrophysics,'' and with partial financial support from the Russian Fund for Fundamental Research (Grant No. 98-02-18040) and the ''Kosmion'' Scientific Educational Center.

*)E-mail: chervon@themp.univ.simbirsk.su

[1] A. D. Linde, *Particle Physics and Inflationary Cosmology* (Harwood Acad. Publ., Paris–New York, 1990) [Russ. original, Nauka, Moscow, 1990].
[2] S. V. Chervon and V. M. Zhuravlev, Izv. Vyssh. Uchebn. Zaved. Fiz. **8**, 81 (1996).
[3] P. Coles and F. Lucchin, *Cosmology: the Origin and Evolution of Cosmic Structure* (Wiley, Chichester, 1995).
[4] A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[5] R. H. Brandenberger, in *Physics of the Early Universe, Proceedings of the 36th Scottish Universities Summer School in Physics*, edited by J. A. Peacock, A. F. Heavens, and A. T. Davies (1989), p. 281.
[6] V. N. Lukash and I. D. Novikov, in *Observational and Physical Cosmology*, edited by F. Sunchez, M. Collados, and K. Rebolo (Cambridge Univ. Press, 1990).
[7] S. V. Chervon and V. M. Zhuravlev, in *Abstracts of the Reports at the International School-Seminar ''Foundations of Gravitation and Cosmology,'' Odessa* (RGS, Moscow, 1995), p. 67.
[8] S. V. Chervon, *Nonlinear Fields in the Theory of Gravitation and Cosmology* [in Russian] (Izd. Srednevolzhsk. Nauchn. Tsentra, Ulyanovsk, 1997).
[9] S. V. Chervon, V. M. Zhuravlev, and V. K. Shchigolev, Phys. Lett. B **398**, 269 (1997).
[10] S. V. Chervon, Gravitation Cosmol. **3**, 151 (1997).
[11] J. D. Barrow, Phys. Rev. D **49**, 3055 (1994).
[12] R. Maartens, D. R. Taylor, and N. Roussos, Phys. Rev. D **52**, 3358 (1995).
[13] P. Parsons and J. D. Barrow, Class. Quantum Grav. **12**, 1715 (1995).
[14] V. K. Shchigolev, V. M. Zhuravlev, and S. V. Chervon, JETP Lett. **64**, 71 (1996).
[15] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972) [Russ. transl., Mir, Moscow, 1975].

Translated by James S. Wood
Edited by P. Shelnitz

Oscillatory phenomena in cold matter with four-fermion interaction

A. S. Vshivtsev*)

Moscow Institute of Radio Engineering, Electronics, and Automation, 117454 Moscow, Russia

M. A. Vdovichenko

Moscow State (M. V. Lomonosov) University, 117899 Moscow, Russia

K. G. Klimenko†)

Institute of High-Energy Physics, 142284 Protvino, Moscow Region, Russia

(Submitted 29 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 418–436 (August 1998)

The phase structure of the Nambu–Jona-Lasinio model in space–time with the topology $R^3\times S^1$ (spatial coordinate compactified) and chemical potential μ is investigated. Phase portraits of the model are constructed in the $\mu\lambda$ plane ($\lambda=1/L$, where L is the size of the dimension S^1) in the case with periodic boundary conditions. It is shown here for the first time that there exist in the model an infinite number of both chirally symmetric massless phases and asymmetric massive phases, between which (as a rule) transitions are second-order. Because of this phase structure, changes in the parameter λ induce oscillations in the particle density, fermion mass, and the critical curve on which chiral invariance is restored. © *1998 American Institute of Physics.* [S1063-7761(98)00408-9]

1. INTRODUCTION

A study of effects associated with the presence of a finite density, temperature, and external fields is of current interest and can reveal new phenomena that were previously unknown. These effects include oscillations in the magnetization of an electron gas in metals[1] and the prediction of the existence of a ''Yang–Mills'' crystal.[2] Besides, the study of matter under extreme conditions is necessary in order to construct a consistent theory of strong interactions[3] and also to construct various models of stars.[4]

Studies of the action of external factors for nonlinear four-fermion theories have led to the discovery of new effects. This is very important since theories with this kind of interaction find application in the explanation of a wide class of phenomena. Thus, we may point to their use in the explanation of superconductivity[5] and high-temperature superconductivity.[6]

We may also mention the general theoretical interest in theories of this type, motivated by the fact that a unified theory of all of the forces of nature, including the gravitational force, is still far from realized. Therefore it makes sense to examine quantum field theories in the spaces with a nontrivial metric and topology in which the Universe most probably existed in the early stages of its evolution.[7] Here special attention has been given, in particular, to spaces of the form $R^d\times S^1\times\ldots\times S^1$. The point here is that compactification of one or more spatial coordinates occurs in some superstring theories[8] and it is convenient to describe phenomena like the Casimir effect in spaces with such topology;[9] finally, when investigating theories in spaces with the simplest nontrivial topology it is possible to make generalizations and predict properties of physical systems in spaces with more complicated structure.

The concept of dynamical symmetry-breaking, i.e., spontaneous symmetry-breaking without the introduction of Higgs fields, is very fruitful in elementary particle physics. Such a mechanism of chiral symmetry-breaking was first revealed in models with a four-fermion interaction,[10,11] the simplest of which has a Lagrangian of the form

$$L_\psi=\sum_{k=1}^{N}\bar{\psi}_k i\hat{\partial}\psi_k+\frac{G}{2N}\left[\left(\sum_{k=1}^{N}\bar{\psi}_k\psi_k\right)^2+\left(\sum_{k=1}^{N}\bar{\psi}_k i\gamma_5\psi_k\right)^2\right] \tag{1}$$

and in $(3+1)$-dimensional space is called the Nambu–Jona-Lasinio (NJL) model. We will consider this model in the $1/N$ expansion. To realize this program we must choose the N-fermion version of the model. Here it should be pointed out that the expansion in powers of $1/N$, characteristic of models of Nambu–Jona-Lasinio and Gross–Neveu type, is quasi-classical in this parameter, and the parameter $1/N$ is an analog of the Planck constant. Note that for the given type of expansion, at each order in the given expansion parameter, we have a set of infinite series of Feynman diagrams.

A rigorous mathematical treatment of models of this type as quantization of classical mechanics with a nonlinear phase space was given in Refs. 12. From the standpoint of physical applications of this method it is important that despite the strict mathematical requirement that the parameter N tend to infinity, in reality its value can be limited to 3. Significant use has been made of this fact, for example, in quantum chromodynamics.[3,13] The results so obtained (even for this value of N) are nonetheless such that qualitatively they correctly reproduce the properties of the physical sys-

tem described in terms of this model. Moreover, available treatments of the $1/N$ expansion which include higher orders of the expansion indicate that these improvements do not alter the physical picture—they only refine some of the numerical values of the parameters.[14] All of this gives us reason to place some hope in the reliability of the results obtained on the basis of this approach.

In what follows we consider the N-fermion version of the model, which is invariant with respect to simple continuous chiral transformations:

$$\psi_k \to \exp(i\theta\gamma_5)\psi_k \quad (k=1,\ldots,N). \tag{2}$$

The idea of dynamical symmetry-breaking also finds application in the explanation of superconductivity[15] and high-temperature superconductivity,[16] in the description of low-energy meson physics,[17] the construction of alternative models of the electroweak interaction,[18] etc. It was recently shown that one source of dynamical symmetry-breaking (of chiral symmetry, that is) is external magnetic and chromomagnetic fields.[19]

One interesting fact demonstrated recently is that low-energy NJL models of can be consistently constructed from the first principles of quantum chromodynamics (QCD).[20] In this light, the question of predicting the effects that might be revealed by such models acquires some importance. It is also known that the QCD ground state contains a nonzero gluon condensate $\langle 0|F_a^{\mu\nu}F_{\mu\nu}^a|0\rangle = 0.215\ \mathrm{GeV}^2$, whose structure (on the basis of model lattice calculations) has a chromomagnetic component.

As one could expect, by reason of its numerous physical applications the phenomenon of dynamical symmetry-breaking has also been studied in spaces with nontrivial metric and topology.[21–24] In particular, Refs. 22 and 23 considered the properties of various models with a four-fermion interaction in spaces of the form $R^2\times S^1\times S^1$ and $R^1\times S^3$, and Ref. 24 examined the structure of the vacuum in NJL models in the presence of an external magnetic field in a space with nonzero curvature.

Since in reality many phenomena take place under conditions of nonzero particle density (superconductivity, the Hall effect, processes in neutron stars, etc.), in the present paper we also examine the effect of dynamical symmetry-breaking with a chemical potential μ. We consider the phase structure of NJL models in a multiply connected space of the form $R^3\times S^1$ for nonzero values of μ. (Here one of the spatial coordinate axes has been compactified, the length of the dimension S^1 is equal to L, and the fermion fields satisfy periodic boundary conditions.) It turns out that the presence of a chemical potential radically alters the phase portrait of the model and leads to oscillations of the particle number density and also of other physical quantities associated with it. This is one aspect of the given theme. Another, and perhaps more important aspect is the fact that the scope of the given model problem (after some modifications) can be broadened to a larger circle of physically more realistic problems. Such problems include four-fermion models which allow for such interaction with a magnetic field, or fermions in periodic structures (crystalline lattices, etc.).[25] General observed physical effects for these models include oscillations in the phase plane, sets of massive and massless phases, and also oscillations of the particle density in these phases.

As our method for investigating the theory described by the Lagrangian (1) we use the effective potential method.[7,13] As is well known, it is in fact a somewhat different form of representation of the method of ''quasi-means.'' Therefore, the results obtained using it can be rigorously justified, for example, in terms of functional integrals.

2. PHASE STRUCTURE OF A MODEL WITH $\mu\neq 0$, $L=\infty$ AND WITH $\mu=0$, $L\neq\infty$

On the conceptual plane, the present study is a continuation of a study of the vacuum begun in Refs. 26 and 27 in the NJL model under various external conditions.[1)] We will formulate the main results obtained in Refs. 26 and 27, and derive the necessary mathematical formulas needed in what follows.

2.1. The case $\mu\neq 0$, $L=\infty$

To start with, let us recall the properties of the vacuum in the theory (1) for $\mu=0$ in Minkowski space, i.e., for $L=\infty$. With this goal in mind, instead of expression (1), we consider the auxiliary Lagrangian

$$L_\sigma = \bar\psi i\hat\partial\psi - \bar\psi(\sigma_1 + i\sigma_2\gamma_5)\psi - \frac{N}{2G}(\sigma_1^2+\sigma_2^2) \tag{3}$$

(here for simplicity we have dropped the subscript k labeling the Fermi fields), which in the equation of motion for the auxiliary boson fields $\sigma_{1,2}$ is equivalent to the original Lagrangian (1).

The effective action of the model to leading order in $1/N$ is given by the following expression:

$$\exp[iNS_{\mathrm{eff}}(\sigma_{1,2})] = \int D\bar\psi D\psi \exp\left(i\int L_\sigma d^4x\right),$$

where

$$S_{\mathrm{eff}}(\sigma_{1,2}) = -\int d^4x \frac{\sigma_1^2+\sigma_2^2}{2G} - i\ln\det(i\hat\partial - \sigma_1 - i\gamma_5\sigma_2). \tag{4}$$

Taking the fields $\sigma_{1,2}$ to be independent of the space–time coordinates (because we seek the vacuum solutions), we have by definition

$$S_{\mathrm{eff}}(\sigma_{1,2}) = -V_{\mathrm{eff}}(\sigma_{1,2})\int d^4x, \tag{5}$$

where

$$V_{\mathrm{eff}}(\sigma_{1,2}) = \frac{\Sigma^2}{2G} + 2i\int\frac{d^4p}{(2\pi)^4}\ln(\Sigma^2-p^2) \equiv V_0(\Sigma), \tag{6}$$

$$\Sigma = \sqrt{\sigma_1^2+\sigma_2^2}.$$

Transforming in expression (6) to the Euclidean metric ($p_0\to ip_0$) and introducing the Lorentz-invariant cutoff of the range of integration ($p^2\leqslant\Lambda^2$), we obtain

$$V_0(\Sigma)=\frac{\Sigma^2}{2G}-\frac{1}{16\pi^2}\left\{\Lambda^4\ln\left(1+\frac{\Sigma^2}{\Lambda^2}\right)+\Lambda^2\Sigma^2-\Sigma^4\ln\left(1+\frac{\Lambda^2}{\Sigma^2}\right)\right\}. \quad (7)$$

The condition that the function (7) be stationary has the form

$$\frac{\partial V_0(\Sigma)}{\partial\Sigma}=0=\frac{\Sigma}{4\pi^2}\left\{\frac{4\pi^2}{G}-\Lambda^2+\Sigma^2\ln\left(1+\frac{\Lambda^2}{\Sigma^2}\right)\right\}. \quad (8)$$

Hence it is clear that for $G<G_c=4\pi^2/\Lambda^2$ Eq. (8) has no solutions besides $\Sigma=0$, i.e., in this case the fermions are massless and the chiral invariance (2) is not violated.

For $G>G_c$ the stationary condition (8) will have one nontrivial solution, containing the global minimum of the potential $V_0(\Sigma)$, which implies spontaneous breaking of chiral symmetry and the appearance on the fermions of nonzero mass.

Let us now assume that $\mu>0$. This case was considered in detail in Ref. 26, which also derived an expression for the corresponding effective potential [the form of $V_0(\Sigma)$ is given by expression (7)]:

$$V_\mu(\Sigma)=V_0(\Sigma)-2\int\frac{d^3p}{(2\pi)^3}\theta(\mu-\sqrt{\Sigma^2+p^2})\times(\mu-\sqrt{\Sigma^2+p^2}). \quad (9)$$

Here $\theta(x)$ is the Heaviside step function. Calculating the integral appearing in this formula, we have

$$V_\mu(\Sigma)=V_0(\Sigma)-\frac{\theta(\mu-\Sigma)}{16\pi^2}\left\{\frac{10}{3}\mu(\mu^2-\Sigma^2)^{3/2}-2\mu^3\sqrt{\mu^2-\Sigma^2}+\Sigma^4\ln\left[\frac{(\mu+\sqrt{\mu^2-\Sigma^2})^2}{\Sigma^2}\right]\right\}. \quad (10)$$

A study of the absolute minimum of this potential[26] revealed previously unknown properties of the NJL model. In particular, it was shown there that in this case for nonzero values of the chemical potential the state with massive fermions is described by two different phases, the transition between which is second-order. It was also shown there that the chiral symmetry of the model can be restored with the help of both first-order and second-order phase transitions depending on the values of the model parameters. A phase portrait of the NJL model containing two tricritical points was constructed in the (μ,M) plane, where M is the dynamic mass of the fermions, for $\mu=0$.

2.2. $\mu=0$, $L\neq\infty$

We now assume that space–time has the structure $R^3\times S^1$, where one of the spatial coordinates has been compactified and the length of the dimension S^1 is a finite value, L. Here we must distinguish two cases. In the first of these the fields of the model (1) satisfy periodic boundary conditions in the compactified coordinate, i.e.,

$$\psi(t,x+L,y,z)=\psi(t,x,y,z). \quad (11)$$

In the second, the fields are antiperiodic in this coordinate. However, this latter case will not be considered in this paper.

Now, in order to obtain the effective potential for $L=\infty$, it is necessary to replace the integral over p_1 in expression (6) by a sum over the discrete values p_{1n} according to the following rule:

$$\int\frac{dp_1}{2\pi}f(p_1)\to\frac{1}{L}\sum_{n=-\infty}^{\infty}f(p_{1n}),\qquad p_{1n}=\frac{2\pi n}{L},\quad n=0,\pm1,\pm2,\ldots. \quad (12)$$

Transforming in the newly obtained expression to the Euclidean metric and summing over n (Ref. 30), we find the effective potential of the NJL model with topology $R^3\times S^1$:

$$V_L(\Sigma)=V_0(\Sigma)-\frac{2}{\pi^2L}\int_0^\infty dx\,x^2\ln[1-\exp(-L\sqrt{x^2+\Sigma^2})]. \quad (13)$$

The potential (13) has an ultraviolet divergence which is concentrated in the first term. To eliminate it, we will, as before, regularize with the help of a Lorentz-invariant cutoff of the range of integration in expression (6). This leads to expression (7) for $V_0(\Sigma)$. We will examine the absolute minimum of the function (13) in the variable Σ. Here, by virtue of the parity of $V_L(\Sigma)$, we may consider only nonnegative values of Σ. The condition that this function be stationary has the form

$$\frac{\partial V_L(\Sigma)}{\partial\Sigma}=0=\frac{2\Sigma}{\pi^2}\{F(\Sigma)-I(\Sigma)\}, \quad (14)$$

where

$$F(\Sigma)=\frac{\pi^2}{2G}-\frac{\Lambda^2}{8}+\frac{\Sigma^2}{8}\ln\left[1+\frac{\Lambda^2}{\Sigma^2}\right],\qquad I(\Sigma)=\int_0^\infty\frac{x^2dx}{\sqrt{x^2+\Sigma^2}}\frac{1}{\exp(L\sqrt{x^2+\Sigma^2})-1}. \quad (15)$$

Hence (see Ref. 27) it follows that

$$F(0)=\frac{\pi^2}{2G}-\frac{\Lambda^2}{8},\quad I(0)=\frac{\pi^2}{6L^2}. \quad (16)$$

It is not hard to see that on the half-interval $\Sigma\geqslant0$ the function $F(\Sigma)$ grows monotonically from the value $F(0)$ at $\Sigma=0$ to $\pi^2/(2G)$ at $\Sigma=\infty$. On the other hand, $I(\Sigma)$ is a monotonically decreasing function, decreasing from $I(0)$ to zero. Consequently, the graphs of this function for $F(0)<I(0)$ necessarily intersect at a single point $\Sigma_0(L)$, which is a nontrivial solution of the stationary condition (14). We introduce the new variable $\lambda=1/L$. Then (see Ref. 27) the nonzero root $\Sigma_0(\lambda)$ $(\equiv\Sigma_0(L))$ of Eq. (14) grows monotonically as a function of λ.

We consider two cases: 1) $G>G_c$, 2) $G<G_c$ [the quantity G_c is defined following Eq. (8)]. In the first case, as

follows from Eq. (16), we have $F(0)<0$, so for all values of λ Eq. (14) has a nontrivial solution $\Sigma_0(\lambda)$, for which the potential $V_L(\Sigma)$ has an absolute minimum. Consequently, for $G>G_c$ the vacuum of the NJL model does not have chiral symmetry. Besides, it is obvious that $\Sigma_0(\lambda)\to M$ as $\lambda\to 0$, where M is the solution of the equation of stationarity for $L=\infty$.

Let us now consider the second case, for which $F(0)>0$. We introduce the notation

$$F(0)=\frac{\pi^2}{2G}-\frac{\Lambda^2}{8}\equiv\frac{\pi^2}{6}\lambda_0^2. \tag{17}$$

It is clear that for $\lambda<\lambda_0$ (i.e., for $F(0)>I(0)$) the equation of stationarity (14) has only a trivial solution, and the symmetry of the NJL model is not broken. At the point $\lambda=\lambda_0$ a second-order phase transition takes place in the model, and for $\lambda>\lambda_0$ the symmetry of the model is broken spontaneously since in this case the global minimum point of the potential, $\Sigma_0(\lambda)$, is not equal to zero. It is not hard to obtain the asymptotic limit of this solution of the equation of stationarity as $\lambda\to\lambda_{0^+}$ (Ref. 27):

$$\Sigma_0(\lambda)=\frac{2}{3}\pi(\lambda-\lambda_0)+o(\lambda-\lambda_0). \tag{18}$$

It can also be shown[27] that both for $G>G_c$ and for $G<G_c$ the solution $\Sigma_0(\lambda)$ has the following asymptotic limit as $\lambda\to\infty$:

$$\Sigma_0(\lambda)\sim 2\pi\lambda\cdot(2.719\ldots). \tag{19}$$

Finally, for what follows note should be made of the obvious fact that $F(\Sigma)-I(\Sigma)$ is a monotonically increasing function on the half-interval $\Sigma\geqslant 0$.

3. PHASE STRUCTURE OF THE MODEL FOR $\mu\neq 0$, $L\neq\infty$ IN THE CASE $G<G_C$

The effective potential of the NJL model to leading order in $1/N$ in this case can be found from Eq. (9) with the help of the transformation (12) [we will be dealing here only with periodic boundary conditions (11)]:

$$V_{\mu L}(\Sigma)=V_L(\Sigma)-\frac{\lambda}{6\pi}\sum_{n=0}^{\infty}\alpha_n\theta(\mu-\sqrt{\Sigma^2+(2\pi\lambda n)^2})\times[\mu-\sqrt{\Sigma^2+(2\pi\lambda n)^2}]^2[\mu+2\sqrt{\Sigma^2+(2\pi\lambda n)^2}]. \tag{20}$$

Recall that here $\lambda=1/L$. In addition, $\alpha_n=2-\delta_{n0}$. The condition that the function (20) be stationary has the form

$$\frac{\partial V_{\mu L}(\Sigma)}{\partial\Sigma}=\frac{2\Sigma}{\pi^2}\left\{F(\Sigma)-I(\Sigma)+\frac{\lambda\pi}{2}\times\sum_{n=0}^{\infty}\alpha_n\theta(\mu-\sqrt{\Sigma^2+(2\pi\lambda n)^2})\times[\mu-\sqrt{\Sigma^2+(2\pi\lambda n)^2}]\right\}\equiv\frac{2\Sigma}{\pi^2}\phi(\Sigma)=0. \tag{21}$$

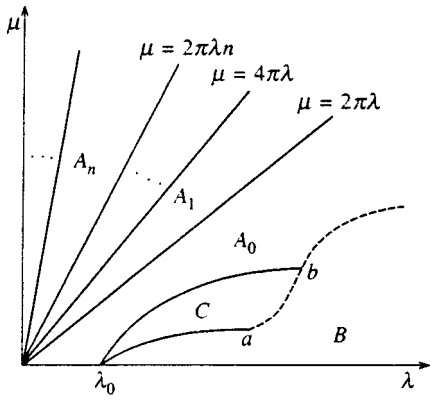


FIG. 1.

We are interested in the phase structure of the model (1); therefore, some phase of the NJL theory must be put in correspondence with each point of the $\mu\lambda$ plane, where $\mu\geqslant 0$ and $\lambda\geqslant 0$. The structure of the potential (20) dictates the following strategy for studying the absolute minimum of this function. We partition the $\mu\lambda$ plane into regions ω_k such that

$$(\mu,\lambda)=\cup_{k=0}^{\infty}\omega_k,\quad \omega_k=\{(\mu,\lambda):2\pi\lambda k\leqslant\mu<2\pi\lambda(k+1)\}. \tag{22}$$

If we fix μ and λ in some arbitrary region ω_k, then only the $k+1$ first terms of the sums in expressions (20) and (21) make a nonzero contribution to these expressions. Below we will run through all these regions in succession, starting with ω_0, and identify the phase structure of the model in each of them.

In the region ω_0 the phase portrait of the model was obtained in Ref. 27 with the help of a quite lengthy analysis of the effective potential (20). It turns out that here for $\Lambda^2<\Lambda_0^2\equiv 16\pi^2\lambda_0^2\cdot(0.926\ldots)$ two phases exist: the chirally symmetric phase A_0 with massless fermions and the phase B, in which the symmetry (2) is spontaneously broken and the single-particle fermion excitations of the vacuum have nonzero mass. If $\Lambda^2>\Lambda_0^2$ holds, then another massive phase C appears in the phase diagram of the model. Just this case is reflected in Fig. 1, where the region ω_0 includes all those points lying below the line $\mu=2\pi\lambda$. In this figure the dashed lines are first-order phase curves, the remaining lines are the critical curves of second-order phase transitions. The points a and b in the region ω_0 are tricritical points.

We now assume that we find ourselves near the boundary of the region ω_0, separating it from ω_1, i.e., in the subregion labeled A_0 in Fig. 1. In this case from the entire sum only the $n=0$ term contributes to the effective potential (20), and its global minimum is located at the point $\Sigma=0$ (see Ref. 27). As we go to the region ω_1, an additional nonzero term appears in the effective potential (20), namely the $n=1$ term. It gives a nonzero contribution to the potential only in a neighborhood of the origin defined by the condition $0\leqslant\Sigma\leqslant\sqrt{\mu^2-(2\pi\lambda)^2}$, and increases the depth of the potential well at the point $\Sigma=0$ still further. Consequently, the point $\Sigma=0$ is also the absolute minimum point of the

potential in the region ω_1. By analogy it is not hard to see that $\Sigma=0$ is the global minimum point of the potential (20) in all the other regions: ω_2, ω_3, etc.

Recall that the thermodynamic potential Ω of the system can be defined as the value of its effective potential at its global minimum. Therefore, in the entire $\mu\lambda$ plane, except for regions B and C in Fig. 1, the thermodynamic potential has the form

$$\Omega(\mu)\equiv V_{\mu L}(0)=V_L(0)-\frac{\lambda}{6\pi}\sum_{n=0}^{\infty}\alpha_n\theta(\mu-2\pi\lambda n)\times(\mu-2\pi\lambda n)^2(\mu+4\pi\lambda n). \tag{23}$$

As is well known, at the point of a first-order (second-order) phase transition all derivatives of the thermodynamic potential, starting with the first (second), are discontinuous. Using this criterion of a phase transition, let us investigate the phase structure of the model in the regions $\omega_1,\omega_2,\ldots$

From Eq. (23) it can be seen that in region A_0, which lies in ω_0, the thermodynamic potential of the system has the form

$$\Omega(\mu)|_{A_0}\equiv V_{\mu L}(0)|_{\omega_0}=V_L(0)-\frac{\lambda\mu^3}{6\pi}. \tag{24}$$

In the region ω_1 it is equal to

$$\Omega(\mu)|_{\omega_1}\equiv V_{\mu L}(0)|_{\omega_1}=\Omega(\mu)|_{A_0}-\frac{\lambda}{3\pi}(\mu-2\pi\lambda)^2(\mu+4\pi\lambda). \tag{25}$$

It is easy to see that like the functions (24) and (25) themselves, their derivatives with respect to μ are discontinuous on the line $\mu=2\pi\lambda$ separating the regions ω_0 and ω_1. However, their second derivatives with respect to μ on this line do not coincide. Consequently, here the second derivative of the thermodynamic potential with respect to μ is discontinuous, and we have a second-order phase transition from the phase A_0 to a different massless phase of the model, A_1. Using similar arguments it is not hard to show that each region ω_k (22) corresponds to a distinct massless phase A_k of the model. The boundaries of these regions are indeed critical curves of second-order phase transitions. A phase portrait of the model for $G<G_c$ and $\Lambda^2>\Lambda_0^2$ is given in Fig. 1.

Let us now assume that $\mu=\text{const}$. In the $\mu\lambda$ plane this corresponds to a straight line intersecting the boundaries of the regions $\omega_0,\omega_1,\ldots,\omega_n,\ldots$ at the points

$$\lambda_0=\frac{\mu}{2\pi},\quad \lambda_1=\frac{\mu}{4\pi},\ldots,\quad \lambda_n=\frac{\mu}{2\pi(n+1)},\ldots$$

respectively. Let us consider the particle number density $N=-\partial\Omega/\partial\mu$ of the system as a function of the parameter λ for $\mu=\text{const}$, i.e., $N=N(\lambda)$. Employing Eq. (23), it is not hard to show that the function $\partial^2\Omega(\mu)/\partial\mu\partial\lambda$ is discontinuous at the boundaries of the regions ω_k. Consequently, the function $N(\lambda)$ is discontinuous for $\lambda\geqslant 0$, and its graph has characteristic kinks at an infinite number of points

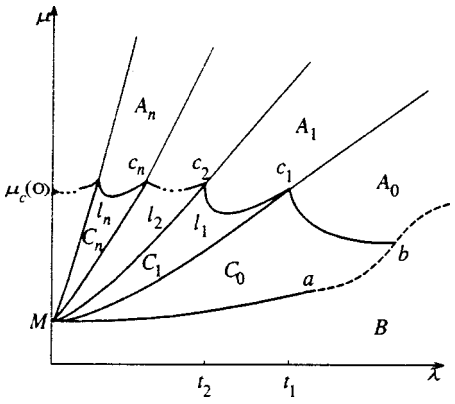


FIG. 2.

$\lambda_0,\lambda_1,\ldots$ (such behavior of a function in practice is called oscillatory), i.e., the fermion number density oscillates as a function of λ. These oscillations are especially clearly pronounced for small (large) values of λ (L). To show this, we apply Poisson's formula[31,32] to the sum in expression (23)

$$\sum_{n=0}^{\infty}\alpha_n\Phi(n)=2\sum_{k=0}^{\infty}\alpha_k\int_0^{\infty}\Phi(x)\cos(2\pi kx)dx, \tag{26}$$

the application of which gives the following result:

$$\Omega(\mu)=V_L(0)-\frac{\mu^4}{12\pi^2}-\sum_{n=0}^{\infty}\left[\frac{4\lambda^4}{\pi^2k^4}-\frac{2\mu\lambda^3}{\pi^2k^3}\times\sin(\mu kL)-\frac{4\lambda^4}{\pi^2k^4}\cos(\mu kL)\right]. \tag{27}$$

Recall that $\lambda=1/L$. It can be seen from Eq. (27) that the thermodynamic potential contains a component that oscillates as a function of the parameter L with frequency $\mu/(2\pi)$. Consequently, the particle number density $N(\lambda)$ also oscillates as a function of the parameter L with the same frequency.

4. PHASE STRUCTURE OF THE MODEL FOR $\mu\neq 0$, $L\neq\infty$ IN THE CASE $G_c<G$

In the case under consideration we have $F(0)<0$ [see Eq. (16)]; therefore, it is convenient to introduce the following notation:

$$F(0)=\frac{\pi^2}{2G}-\frac{\Lambda^2}{8}\equiv-\frac{\pi^2}{6}\bar{\lambda}_0^2. \tag{28}$$

The effective potential and the stationary condition, as before, have the form (20) and (21), respectively. In Ref. 27 the phase structure of the model for $G_c<G$ was investigated in detail in the region ω_0. Results of this analysis for values of the coupling constant in the interval

$$G_c<G<(1.225\ldots)G_c \tag{29}$$

are shown in Fig. 2, where under the straight line $\mu=2\pi\lambda$ (the boundary of the region ω_0) is located a phase portrait of the NJL model for the points $(\mu,\lambda)\in\omega_0$. In this figure the

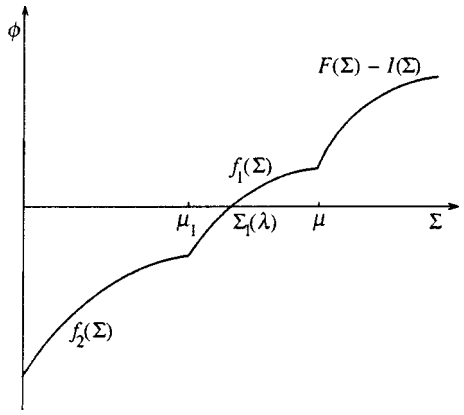


FIG. 3.

solid lines are the critical curves of second-order phase transitions, and the dashed lines are the critical curves of first-order phase transitions. It can be seen that, as for $G<G_c$, the model has two tricritical points, a and b, in the region ω_0. Also note that the curve $c_1 b$ in this figure corresponds to the equation $\phi(0)=0$, and the projection of the point c_1 on the λ axis is $t_1 \equiv \overline{\lambda}_0/\sqrt{5}$. In addition, the curve Ma has the form $\mu=\Sigma_0(\lambda)$, where $\Sigma_0(\lambda)$ is the root of the stationary condition (14) which is the fermion mass in the phase B. (A more detailed description of the phase portrait of the model in the region ω_0 can be found in Ref. 27).

In this section we examine the properties of the vacuum of the NJL model in the regions $\omega_1, \omega_2, \ldots$ In order not to encumber the article with additional calculations, we limit the discussion here to the range of values of the coupling constants bounded by the inequalities (29).

4.1. Particular case

Let us assume, to start with, that $(\mu,\lambda) \in \omega_1$ (i.e., $2\pi\lambda < \mu < 4\pi\lambda$), and the value of λ is fixed ''near'' t_1 so that $\lambda < t_1$. In this case the function $\phi(\Sigma)$ in the stationary condition (21) has the form

$$\phi(\Sigma)=\begin{cases} F(\Sigma)-I(\Sigma), & \text{if} \quad \mu<\Sigma, \\ f_1(\Sigma), & \text{if} \quad \mu_1<\Sigma<\mu, \\ f_2(\Sigma), & \text{if} \quad 0<\Sigma<\mu_1, \end{cases} \tag{30}$$

where

$$\mu_1=\sqrt{\mu^2-(2\pi\lambda)^2}, \quad f_1(\Sigma)\equiv F(\Sigma)-I(\Sigma)+\frac{\lambda\pi}{2}(\mu-\Sigma),$$

$$f_2(\Sigma)\equiv F(\Sigma)-I(\Sigma)+\frac{\lambda\pi}{2}(\mu-\Sigma) + \lambda\pi(\mu-\sqrt{\Sigma^2+(2\pi\lambda)^2}).$$

For the range of values of the coupling constants bounded by inequalities (29) the function $\phi(\Sigma)$ (30) grows monotonically for $\Sigma \geqslant 0$. Its graph is plotted in Fig. 3 for $\mu<2\pi\lambda$, i.e., in the case when $f_1(\mu_1)<0$ and $f_1(\mu)>0$. Thus it is clear that the only zero of this function $\Sigma_1(\mu,\lambda)$ (this zero being the fermion mass in the phase C_0 of the theory; see Fig. 1) is defined by the equation $f_1(\Sigma)=0$. It is important to note that there is a kink in the graph of the function $\phi(\Sigma)$ at the point μ_1. It is also important to note that f_1 and f_2 grow monotonically with increase of the parameter μ. Therefore, as μ is increased, at some value $\mu=\overline{\mu}_1(\lambda)$ this kink reaches the Σ axis. And for $\mu>\overline{\mu}_1(\lambda)$ the point $\Sigma_2(\mu,\lambda)$, being a root of the equation $f_2(\Sigma)=0$, becomes a solution of the stationary condition. Finally, for the value $\mu=\mu_c(\lambda)$ defined by the condition $\phi(0)=f_2(0)=0$ the fermion mass vanishes and a transition to the chirally symmetric phase of the model takes place. Employing definition (30), we can show that

$$\frac{\mu_c(\lambda)}{\pi}=\frac{13\lambda^2+\overline{\lambda}_0^2}{9\lambda}. \tag{31}$$

Note that the situation described above is valid for values $t_2<\lambda<t_1$, where t_2 is the intersection point of curve (31) with the upper boundary of the region ω_1: $t_2=\overline{\lambda}_0/\sqrt{23}$.

We will now show that a second-order phase transition takes place at the point $\mu=\overline{\mu}_1(\lambda)$. To this end, note that for $\mu<\overline{\mu}_1(\lambda)$ the thermodynamic potential of the system $\Omega(\mu)$ has the form

$$\Omega(\mu)\big|_{\mu<\overline{\mu}_1(\lambda)} \equiv \Omega_0(\mu)=V_{\mu L}(\Sigma_1)=V_L(\Sigma_1) - \frac{\lambda}{6\pi}(\mu-\Sigma_1)^2(\mu+2\Sigma_1), \tag{32}$$

and for $\mu_c(\lambda)>\mu>\overline{\mu}_1(\lambda)$ it is given by

$$\Omega(\mu)\big|_{\mu>\overline{\mu}_1(\lambda)} \equiv \Omega_1(\mu)=V_{\mu L}(\Sigma_2) = V_L(\Sigma_2)-\frac{\lambda}{6\pi}(\mu-\Sigma_2)^2(\mu+2\Sigma_2) - \frac{\lambda}{3\pi}[\mu-\sqrt{\Sigma_2^2+(2\pi\lambda)^2}]^2 \times[\mu+2\sqrt{\Sigma_2^2+(2\pi\lambda)^2}]. \tag{33}$$

In Eqs. (32) and (33) we have introduced the abbreviated notation $\Sigma_{1,2}\equiv\Sigma_{1,2}(\mu,\lambda)$. Also, it is necessary to bear in mind that in the case under consideration only those terms in the effective potential (20) corresponding to $n\leqslant 1$ and satisfying the inequalities $\Sigma_2\leqslant\mu_1\leqslant\Sigma_1\leqslant\mu$ are nonzero.

Differentiating Ω_0 with respect to μ, we have

$$\frac{d\Omega_0(\mu)}{d\mu}=\left\{\frac{\partial V_{\mu L}(\Sigma)}{\partial\mu}+\frac{\partial V_{\mu L}(\Sigma)}{\partial\Sigma}\frac{\partial\Sigma}{\partial\mu}\right\}\Bigg|_{\Sigma=\Sigma_1(\mu,\lambda)}. \tag{34}$$

Since Σ_1 is a solution of the stationary condition, the second term on the right-hand side of Eq. (34) vanishes. Taking Eq. (32) into account, we obtain

$$\frac{d\Omega_0(\mu)}{d\mu}=\frac{\partial V_{\mu L}(\Sigma)}{\partial\mu}\Bigg|_{\Sigma=\Sigma_1}=-\frac{\lambda}{2\pi}[\mu^2-\Sigma_1^2(\mu,\lambda)]. \tag{35}$$

Analogously, we obtain

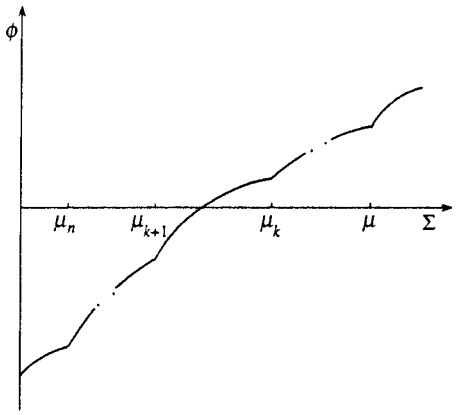


FIG. 4.

$$\frac{d\Omega_1(\mu)}{d\mu}=-\frac{\lambda}{2\pi}[\mu^2-\Sigma_2^2(\mu,\lambda)]$$

$$-\frac{\lambda}{\pi}[\mu^2-\Sigma_2^2(\mu,\lambda)-(2\pi\lambda)^2]. \tag{36}$$

Below we will have need of relations which follow in a transparent way from Fig. 3 (see above):

$$\Sigma_1(\overline{\mu}_1(\lambda),\lambda)=\Sigma_2(\overline{\mu}_1(\lambda),\lambda)=\sqrt{\overline{\mu}_1^2(\lambda)-(2\pi\lambda)^2}. \tag{37}$$

Thanks to relation (37) it is not hard to see that the values of the functions $\Omega_{0,1}(\mu)$, and also their derivatives with respect to μ, coincide at the point $\mu=\overline{\mu}_1(\lambda)$. However, if we differentiate expressions (35) and (36) once more with respect to μ (in this case we need expressions of the form $\partial\Sigma_{1,2}/\partial\mu$, which can be found without any difficulty from the stationary condition), it turns out that the values of the second derivatives of the functions $\Omega_{0,1}$ do not coincide at this point (similar calculations were carried out in Ref. 26). Thus, the second derivative of the thermodynamic potential $\Omega(\mu)$ with respect to μ is discontinuous at the point $\mu=\overline{\mu}_1(\lambda)$. Consequently, this point is the critical point of a second-order phase transition (recall that the value of the parameter λ is fixed). If we allow λ to vary, then the function $\mu=\overline{\mu}_1(\lambda)$ in the $\mu\lambda$ plane will correspond to the critical curves of a set of second-order phase transitions, l_1, which separate the two massive phases of the system, C_0 and C_1 (see Fig. 2). Finally, note that the curve l_1 is given implicitly by the equation [this follows from Eq. (37)]

$$\phi(\sqrt{\mu^2-(2\pi\lambda)^2})=0. \tag{38}$$

4.2. The general case

For the general case, when the parameter λ is fixed without any restrictions in the interval $0<\lambda<t_1$, the graph of the function $\phi(\Sigma)$ for $(\mu,\lambda)\in\omega_n$ is given in Fig. 4, where

$$\mu_k=\sqrt{\mu^2-(2k\pi\lambda)^2},\quad k=1,2,\ldots\ . \tag{39}$$

If the coupling constant is fixed in the interval (29), then $\phi(\Sigma)$ grows monotonically for $0\leqslant\Sigma$ for fixed values of μ. In addition, for fixed values of Σ the function $\phi(\Sigma)$ grows monotonically as a function of the parameter μ. It can be seen from Fig. 4 that the first derivative of this function is discontinuous at the points μ_k, i.e., that these points are kinks of the graph of this function. With growth of μ these kinks are shifted upward and to the right, and as one enters into the region ω_{n+1} another kink appears in the graph of the function at the point μ_{n+1}. At the same time, the nontrivial root of the stationary condition (21) is shifted to the left, and as the next kink crosses the Σ axis a second-order phase transition from one massive phase to another takes place in the theory (this can be shown in a way analogous to the case considered above). In the $\mu\lambda$ plane the critical curves $\mu=\overline{\mu}_k(\lambda)$, where $k=1,2,\ldots$ correspond to these phase transitions (in Fig. 2 these curves are the lines $l_1,l_2,\ldots$, which separate the massive phases $C_0,C_1,C_2,\ldots$, one from the other). Each critical curve l_k, obviously, is defined by the equation

$$\phi(\mu_k)\equiv\phi(\sqrt{\mu^2-(2k\pi\lambda)^2})=0. \tag{40}$$

Finally, for those values of the chemical potential $\mu=\mu_c(\lambda)$ that are assigned implicitly by the equation

$$\phi(0)=0=-\frac{\pi^2}{6}\overline{\lambda}_0^2-\frac{\pi^2}{6}\lambda^2+\frac{\lambda\pi}{2}$$

$$\times\sum_{n=0}^{\infty}\alpha_n\theta(\mu-2\pi\lambda n)(\mu-2\pi\lambda n), \tag{41}$$

the nontrivial root of the stationary condition (identically equal to the fermion mass) vanishes. The dependence $\mu_c(\lambda)$ is depicted in Fig. 2 by the critical curves of the second-order phase transitions $\mu_c(0)c_nc_2c_1b$, on which the chiral symmetry of the model is restored. (In the following section the properties of this curve are discussed in more detail.) The curve $\mu_c(\lambda)$ intersects the boundaries of the regions ω_k at the points $c_1,c_2,\ldots,c_n,\ldots$. The lines $l_1,l_2,\ldots,l_n,\ldots$, which all begin at the point with coordinates $(M,0)$ in Fig. 2, terminate at these same points. Thus, it follows from our analysis that infinitely many massive phases $C_0,C_1,C_2,\ldots$ exist in the NJL model (see Fig. 2).

Analogous to the case $G<G_0$, it is possible to show without any effort that infinitely many massless phases $A_0,A_1,\ldots,A_n,\ldots$ exist above the curve $\mu_c(\lambda)$ (see Fig. 2), the boundaries between which pass along the boundaries of the regions ω_n, i.e., along the lines $\mu=2\pi n\lambda$, where $n=1,2,\ldots$ Each of the lines $\mu=2\pi n\lambda$ is a critical curve of a second-order phase transition from one massless phase to another by virtue of the fact that here the thermodynamic potential $\Omega(\mu)$ has a discontinuous second derivative with respect to μ.

5. OSCILLATORY PHENOMENA FOR $G_c<G<(1.225\ldots)G_c$

We will now show that the presence in the NJL model of infinite sets of both massless and massive phases leads to various oscillatory phenomena. This question was already considered in part in Sec. 3, where we showed that in the case $G<G_c$ the particle density of the vacuum of the system oscillates.

We say that a physical quantity oscillates as a function of λ (or any other parameter) if its graph has an infinite number of kinks $\lambda_1,\lambda_2,\ldots$, i.e., its derivative at these points is discontinuous.

Of course, smoother functions can also oscillate; therefore, the above definition requires some clarification. The point is that at zero temperature and in the presence of a chemical potential oscillations of physical quantities, as a rule, satisfy the definition. This is because the number of filled energy levels of the system, i.e., levels lying below the Fermi surface, depends in a discrete way on the values of the external parameters. As an example, we note that the magnetization $M(H)$ and particle density $N(H)$ of a free fermion gas located in an external magnetic field H at zero temperature have an infinite number of points of discontinuity of the first derivative in H.[31,32] Just this property can be properly called oscillatory behavior of the magnetization and density, which for small values of H can be extracted from the functions $M(H)$ and $N(H)$ explicitly with the help of the Poisson formula (26) (Ref. 33).

The behavior of the thermodynamic potential becomes smoother at nonzero temperature; however, we do not consider this case here, and of the types of oscillations analyzed in the present section are in accord with our definition.

5.1. Oscillations of the critical curve $\bar{\mu}_c(\lambda)$

Recall that the function $\mu_c(\lambda)$ is defined as the solution of Eq. (41). Let us examine its behavior in the region Ω_k $(2\pi k\lambda\leqslant\mu<2\pi(k+1)\lambda)$ (22). In this case, only the terms corresponding to $n\leqslant k+1$ in the sum in Eq. (41) are nonzero. After summing them up, we easily obtain an expression

$$\mu_c(\lambda)|_{\Omega_k}\equiv\mu_{(k)}(\lambda)=\frac{2\pi\{[6k(k+1)+1]\lambda^2+\overline{\lambda}_0^2\}}{6(2k+1)\lambda}, \tag{42}$$

defining the function $\mu_c(\lambda)$ inside the region Ω_k of the $\mu\lambda$ plane. Thus we have

$$\mu_c(\lambda)=\mu_{(k)}(\lambda)\ \text{ for }\ t_{k+1}\leqslant\lambda\leqslant t_k,\quad k=1,2,3,\ldots, \tag{43}$$

where t_k is the value of the parameter λ for which the graph of the function $\mu_{(k)}(\lambda)$ intersects the right-hand boundary of the region Ω_k, i.e., the straight line $\mu=2\pi k\lambda$:

$$t_k=\frac{\overline{\lambda}_0}{\sqrt{6k^2-1}}. \tag{44}$$

It should also be noted that $\mu_{(k)}(t_k)=\mu_{(k-1)}(t_k)$; consequently, the function $\mu_c(\lambda)$ (43) is discontinuous over the entire interval $\lambda>0$. It is easy to show that

$$\left.\frac{d\mu_{(k-1)}(\lambda)}{d\lambda}\right|_{\lambda\to t_{k+}}=\frac{\pi(2-6k)}{3(2k-1)}<0,$$

$$\left.\frac{d\mu_k(\lambda)}{d\lambda}\right|_{\lambda\to t_{k-}}=\frac{\pi(2+6k)}{3(2k+1)}>0.$$

This means that the function $\mu_c(\lambda)$ is not differentiable at an infinite number of points t_k $(k=1,2,\ldots)$ and these points are kinks of this function. Consequently, $\mu_c(\lambda)$ oscillates. In order to extract in explicit form the component of this function that oscillates as $\lambda\to0$ $(L\to\infty)$, we apply the Poisson summation formula (26) to expression (41). After some uncomplicated transformations, Eq. (41) reduces to the form

$$\phi(0)=-\frac{1}{6}+\frac{\mu_c^2(\lambda)}{4\pi^2\overline{\lambda}_0^2}-\frac{\lambda^2}{\pi^2\overline{\lambda}_0^2}\sum_{n=1}^{\infty}\frac{\cos[n\mu_c(\lambda)L/2]}{n^2}\equiv0. \tag{45}$$

Hence it can be seen that $\mu_c(0)=2\pi\overline{\lambda}_0/\sqrt{6}$. In addition, it is possible to obtain the following asymptotic formula for the critical curve $\mu_c(\lambda)$ in the limit $\lambda\to0$:

$$\mu_c(\lambda)\approx\frac{2\pi\overline{\lambda}_0}{\sqrt{6}}\left\{1+\frac{3\lambda^2}{\pi^2\overline{\lambda}_0^2}\sum_{n=1}^{\infty}\frac{\cos(n\pi\overline{\lambda}_0L/\sqrt{6})}{n^2}\right\}. \tag{46}$$

It can be seen that the sum appearing in formula (46) is a periodic function of the parameter L. Thus, $\mu_c(\lambda)$ has an oscillatory component with oscillations whose frequency in the limit $L\to\infty$ is equal to $\overline{\lambda}_0/(2\sqrt{6})$, and whose amplitude falls off as $\lambda\to0$. It is clear from formula (45) that the frequency of the oscillations of this function depends on the value of the parameter λ.

5.2. Oscillations of the fermion condensate

We will now show that the nontrivial solution $\Sigma(\mu,\lambda)$ (identically equal to the fermion mass or fermion condensate) of the stationary condition (21), which, as can be easily gathered from Fig. 4, is the global minimum point of the effective potential, also oscillates with variation of $\lambda=1/L$.

In the phase diagram of Fig. 2 we can mentally draw a straight line $\mu=\text{const}$ such that $M<\mu<\mu_c(0)$. We denote the projections of its intersection points with the critical curves $l_1,l_2,\ldots$ onto the λ axis as $\lambda_1,\lambda_2,\ldots$, respectively. With the help of Eq. (21) it is possible to show that at these points the fermion mass $\Sigma(\mu,\lambda)$, considered only as a function of the parameter λ, has discontinuities in its first derivative (in order not to overload the article with elaborate formulas, we omit these calculations here). Consequently, according to the above definition, the fermion condensate oscillates with variation of λ.

In Ref. 26 it was shown that for the values of the chemical potential considered there the solution $\Sigma(\mu,0)\equiv m(\mu)$ satisfies the equation

$$f(m)\equiv F(m)+\frac{\mu}{4}\sqrt{\mu^2-m^2}-\frac{m^2}{4}\times\ln\left(\frac{\mu+\sqrt{\mu^2-m^2}}{m}\right)=0, \tag{47}$$

where for simplicity we have omitted the dependence on m and μ. We now apply the Poisson summation formula (26) to Eq. (21), which yields

$$0=f(\Sigma)-I(\Sigma)+\sqrt{\mu^2-\Sigma^2}\times\sum_{n=1}^{\infty}\int_0^1[\mu-\sqrt{\Sigma^2+(\mu^2-\Sigma^2)x^2}]\times\cos(n\sqrt{\mu^2-\Sigma^2}Lx)dx. \quad (48)$$

Let us consider Eq. (48) for small values of λ (large L). In this case $\Sigma(\mu,\lambda)\approx m(\mu)$ and $f(\Sigma)\approx f'(m(\mu))[\Sigma-m(\mu)]$ [the latter relation is obvious if we take Eq. (47) into account]. The asymptotic behavior of the integrals in Eq. (48) can be found using the following formula, found after multiple integration by parts:

$$\int g(x)\cos(Lx)dx=\frac{1}{L}g(x)\sin(Lx)+\frac{1}{L^2}g'(x)\cos(Lx)+o\left(\frac{1}{L^2}\right).$$

Finally, noting the exponential smallness of expressions of the type $I(\Sigma)$ as $L\to\infty$ (Ref. 34) we obtain the following equation from Eq. (48), valid for small λ:

$$f'(m(\mu))[\Sigma-m(\mu)]-\frac{\lambda^2\sqrt{\mu^2-m^2(\mu)}}{\mu}\times\sum_{n=1}^{\infty}\frac{\cos[n\sqrt{\mu^2-m^2(\mu)}L]}{n^2}+o(\lambda^2)=0. \quad (49)$$

Solving Eq. (49) for Σ, we find

$$\Sigma(\mu,\lambda)=m(\mu)+\frac{\lambda^2\sqrt{\mu^2-m^2(\mu)}}{\mu f'(m(\mu))}\times\sum_{n=1}^{\infty}\frac{\cos[n\sqrt{\mu^2-m^2(\mu)}L]}{n^2}+o(\lambda^2). \quad (50)$$

5.3. Oscillations of the particle density

In Sec. 3 (the case $G<G_c$) we showed that as the parameter L increases for a fixed value of the chemical potential μ, a theoretically unlimited number of second-order phase transitions from one chirally symmetric, massless phase to another takes place in a quantum system described by the NJL Lagrangian (1). At the instant of each of these phase transitions the derivative with respect to λ of such a physical quantity as the fermion number density $N(L)$ varies discontinuously. From the point of view of an experimenter measuring $N(L)$, the particle density would be seen to oscillate with frequency $\mu/(2\pi)$.

Analogously, in the case $G_c<G<(1.225\ldots)G_c$ for the chemical potential fixed at some value in the interval $\mu_c(0)<\mu$ (see Fig. 2), we have a sequence of phase transitions from one massless phase to another taking place with variation of the parameter λ. Here the particle density oscillates with the same frequency $\mu/(2\pi)$.

Now let us study the question of oscillations of the density for the chemical potential fixed at some value in the interval $M<\mu<\mu_c(0)$. In this case it is obvious that the global minimum point of the potential, $\Sigma(\mu,\lambda)$, is not equal to zero. Therefore, as L increases an infinite number of phase transitions from one massive phase to another is possible in the system. At an interface between phases the derivative of the function $N(L)$ is discontinuous, as a result of which density oscillations are observed. We will describe this process with an analytical expression. First of all, note that the thermodynamic potential of the system is defined as the value of the effective potential at the global minimum point, i.e., by the equality $\Omega(\mu)=V_{\mu L}(\Sigma(\mu,\lambda))$. The particle density in the ground state is

$$N(L)=-\frac{d\Omega_0(\mu)}{d\mu}=-\left\{\frac{\partial V_{\mu L}(\Sigma)}{\partial\mu}+\frac{\partial V_{\mu L}(\Sigma)}{\partial\Sigma}\frac{\partial\Sigma}{\partial\mu}\right\}_{\Sigma=\Sigma(\mu,\lambda)}=\frac{\lambda}{2\pi}\sum_{n=0}^{\infty}\alpha_n\theta(\mu-\sqrt{\Sigma^2(\mu,\lambda)+(2\pi\lambda n)^2})\times[\mu^2-\Sigma^2(\mu,\lambda)-(2\pi\lambda n)^2]. \quad (51)$$

Applying the Poisson summation formula (26) to expression (51), we obtain

$$N(L)=\frac{(\mu^2-\Sigma^2)^{3/2}}{3\pi^2}-\sum_{n=1}^{\infty}\frac{2\lambda^2\sqrt{\mu^2-\Sigma^2}\cos(n\sqrt{\mu^2-\Sigma^2}L)}{\pi^2n^2}+\sum_{n=1}^{\infty}\frac{2\lambda^3\sin(n\sqrt{\mu^2-\Sigma^2}L)}{\pi^2n^3}, \quad (52)$$

where it is understood that Σ is the solution of the stationary condition (21), i.e., it depends on μ and λ and, consequently, at small λ oscillates according to formula (50). With the help of formula (50) it is not hard to extract the leading asymptotic behavior from formula (52) as $\lambda\to0$ for $\mu=\text{const}$:

$$N(L)=\frac{[\mu^2-m^2(\mu)]^{3/2}}{3\pi^2}+\lambda^2\left[\frac{m(\mu)[\mu^2-m^2(\mu)]}{\mu f'(m(\mu))}-2\sqrt{\mu^2-m^2(\mu)}\right]\times\sum_{n=1}^{\infty}\frac{\cos[n\sqrt{\mu^2-m^2(\mu)}L]}{\pi^2n^2}+o(\lambda^2), \quad (53)$$

where $m(\mu)$ is the fermion mass for $\lambda=1/L=0$, which is the solution of Eq. (47).

Thus we see that in the limit $L\to\infty$ the particle density in the ground state of the system oscillates in this parameter in the same way as the fermion condensate, i.e., with frequency $\sqrt{\mu^2-m^2(\mu)}/(2\pi)$.

6. CONCLUSION

In this work we have carried out a detailed study of the phase structure of the Nambu–Jona-Lasinio model in a

space–time with topology $R^3\times S^1$ and chemical potential μ. We have demonstrated here for the first time that an infinite set of massless and massive phases exists in this model. We have indicated the types of phase transitions between these phases. We have demonstrated the possibility of oscillations of the particle number density in this model, and also oscillations of the phase curve. It follows directly from our treatment that for fermions with interaction Lagrangian of the form (1) in a uniform magnetic field oscillatory effects of de Haas–van Alphen type take place,[1,33] and also oscillations in the phase plane $\Sigma\mu$ for the case of finite temperature as well as for cold, dense fermion matter. The existence of such an effect is possible thanks to the special discrete structure of the fermion spectrum in a magnetic field (Landau levels).[1] This same effect is very unusual for the four-fermion model and probably plays a definite role in the physical applications discussed in the Introduction. In our treatment we have assumed the fields σ_1 and σ_2 to be independent of the space–time coordinates. If we reject this requirement and admit the possibility of field fluctuations, then, generally speaking, we arrive at a new physical theory. Such a theory can be handled by summing the entire series in the effective potential method. However, in light of the fact that the oscillations we have found are due to scale parameters of the problem associated with the values of the density, μ, and L, and fluctuations in such a theory are apparently equiprobable, it may be hoped that the revealed effect holds up in more complicated physical situations.

One of us (A. S. V.) would like to express his gratitude to DAAD for making it possible to carry out research on the subject of this paper at the Institute of Physics of Humboldt University (Berlin) and to Prof. Ebert for helpful discussions and for his collaboration.

This work was carried out with the partial financial support of the Russian Fund for Fundamental Research (Grant No. 98-02-16690).

*)E-mail: alexandr@vvas.msk.ru
†)E-mail: kklim@mx.ihep.su

1)Note that early papers[28,29] considered the effect of temperature, the chemical potential, and various external gauge fields on the phase structure of this model.

[1] L. D. Landau, in *Collected Papers of L.D. Landau*, [Ed.] D. ter Haar, Gordon and Breach, New York (1965), Vol. 1 [Russian original, Nauka, Moscow, (1969)]; I. M. Lifshits, *Collected Works* [in Russian] (Nauka, Moscow, 1994).
[2] D. I. Deryagin, D. Yu. Grigoriev, and V. A. Rubakov, Phys. Lett. B **178**, 385 (1986).
[3] A. V. Smilga, Phys. Rep. **291**, 1 (1997); E-prints archive, hep-ph 9612347.
[4] S. Chakrabarty, Phys. Rev. D **54**, 2026 (1996).
[5] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1963).
[6] V. L. Ginzburg and K. G. Maksimov, Sverkhprovodimost': Fiz., Khim., Tekh. **5**, 1543 (1992) [Supercond., Phys. Chem. Technol. **5**, 1505 (1992)].
[7] N. D. Birrell and P. C. W. Davis, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, 1982); I. L. Buchbinder, S. D. Odintsov, and I. L. Shapiro, *Effective Action in Quantum Gravity* (IOP Publishing, Bristol, England, 1992).
[8] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge University Press, Cambridge, 1987).
[9] A. A. Grib, S. G. Mamaev, and V. M. Mostepanenko, *Quantum Vacuum Effects in Strong Fields* [in Russian] (Energoatomizdat, Moscow, 1988); L. H. Ford, Proc. R. Soc. London, Ser. A **368**, 30 (1979); D. J. Toms, Phys. Rev. D **21**, 928 (1980).
[10] Y. Nambu and G. Jona-Lasinio, Phys. Rev. **122**, 345 (1961).
[11] V. G. Vaks and A. I. Larkin, Zh. Eksp. Teor. Fiz. **40**, 282, 1392 (1961) [Sov. Phys. JETP **13**, 192, 979 (1961); B. A. Arbuzov, A. N. Tavkhelidze, and R. N. Faustov, Dokl. Akad. Nauk SSSR **139**, 345 (1961) [Sov. Phys. Dokl. **6**, 598 (1962)].
[12] F. A. Berezin, *Method of Second Quantization* (Academic Press, New York, 1966); F. A. Berezin, Commun. Math. Phys. **63**, 131 (1978).
[13] S. Weinberg, *The Quantum Theory of Fields*, Vols. 1 and 2 (Cambridge Univ. Press, New York, 1996).
[14] G. Gat, A. Kovner, and B. Rosenstein, Phys. Lett. B **240**, 158 (1990).
[15] E. M. Lifshitz and L. P. Pitaevskiĭ, *Statistical Physics, Part 2* (Pergamon Press, Oxford, 1980) [Russian original, Nauka, Moscow, 1978].
[16] G. Semenoff and L. Wijewardhana, Phys. Rev. Lett. **63**, 2633 (1989); N. Dorey and N. Mavromatos, Phys. Lett. B **250**, 107 (1990); A. Kovner and B. Rosenstein, Phys. Rev. B **42**, 4748 (1990); M. Carena, T. E. Clark, and C. E. M. Wagner, Nucl. Phys. B **356**, 117 (1991).
[17] M. K. Volkov, Fiz. Elem. Chastits At. Yadra **17**, 433 (1986) [Sov. J. Part. Nucl. **17**, 186 (1986)]; Fiz. Elem. Chastits At. Yadra **24**, 81 (1993) [Phys. Part. Nuclei **24**, 35 (1993)].
[18] S. Weinberg, Phys. Rev. D **13**, 974 (1976); **56**, 2303 (1997).
[19] K. G. Klimenko, Teor. Mat. Fiz. **89**, 211 (1991); K. G. Klimenko, Z. Phys. C **54**, 323 (1992); K. G. Klimenko, A. S. Vshivtsev, and B. V. Magnitsky, Nuovo Cimento A **107**, 439 (1994); V. P. Gusynin, V. A. Miransky, and I. A. Shovkovy, Phys. Rev. Lett. **73**, 3499 (1994); Phys. Lett. B **349**, 477 (1995).
[20] Yu. A. Simonov, E-prints archive, hep-ph 9704301; 9703205.
[21] T. Hosotani, Phys. Lett. B **126**, 309 (1983); Ann. Phys. **190**, 233 (1989); A. Higuchi and L. Parker, Phys. Rev. D **37**, 2853 (1988); M. Burgess and D. J. Toms, Ann. Phys. **210**, 438 (1991).
[22] T. Inagaki, T. Muta, and S. D. Odintsov, Prog. Theor. Phys. Suppl. **127**, 1 (1997).
[23] T. Inagaki, T. Muta, and S. D. Odintsov, Mod. Phys. Lett. A **8**, 2117 (1993); E. Elizalde, S. Leseduarte, and S. D. Odintsov, Phys. Rev. D **49**, 5551 (1994); Phys. Lett. B **347**, 33 (1995); D. K. Kim and I. G. Koh, Phys. Rev. D **51**, 4573 (1995); H. Forkel, Phys. Lett. B **280**, 5 (1992); Nucl. Phys. A **581**, 557 (1995).
[24] D. M. Gitman, S. D. Odintsov, and Yu. Shil'nov, Phys. Rev. D **54**, 2968 (1996); B. Geyer, L. N. Granda, and S. D. Odintsov, Mod. Phys. Lett. A **11**, 2053 (1996).
[25] V. N. Popov, *Functional Integrals in Quantum Field Theory and Statistical Physics* (Reidel, Dordrecht, 1983).
[26] A. S. Vshivtsev and K. G. Klimenko, JETP Lett. **64**, 245 (1996); A. S. Vshivtsev, V. Ch. Zhukovskiĭ, and K. G. Klimenko, Zh. Éksp. Teor. Fiz. **111**, 1921 (1997) [JETP **84**, 1047 (1997)].
[27] A. S. Vshivtsev, A. K. Klimenko, and K. G. Klimenko, Yad. Fiz. **61**(3), 1 (1998); M. A. Vdovichenko, A. S. Vshivtsev, and K. G. Klimenko, Preprint IFVÉ 97-59 [Inst. High-Energy Phys.], Protvino (1997).
[28] S. Kawati and H. Miyata, Phys. Rev. D **23**, 3010 (1981); J. Fuchs, Z. Phys. C **22**, 83 (1984); V. Bernard, U.-G. Meissner, and I. Zahed, Phys. Rev. D **36**, 819 (1987); Chr. V. Christov and K. Goeke, Acta Phys. Pol. B **22**, 187 (1991); D. Ebert, Yu. L. Kalinovsky, L. Münchow, and M. K. Volkov, Int. J. Mod. Phys. A **8**, 1295 (1993).
[29] D. Ebert and M. K. Volkov, Phys. Lett. B **272**, 86 (1991); S. P. Klevansky and R. H. Lemmer, Phys. Rev. D **39**, 3478 (1989); I. A. Shovkovy and V. M. Turkowski, Phys. Lett. B **367**, 213 (1995).
[30] L. Dolan and R. Jackiw, Phys. Rev. D **9**, 3320 (1974).
[31] L. D. Landau and E. M. Lifshitz, in *Statistical Physics*, Part 1, 3rd ed. (Pergamon Press, Oxford, 1980) [Russian original, Nauka. Moscow, (1976)].
[32] Yu. B. Rumer and M. Sh. Ryvkin, *Thermodynamics, Statistical Physics, and Kinetics* [in Russian] (Nauka, Moscow, 1977); D. Persson and V. Zeitlin, Phys. Rev. D **51**, 2026 (1995); J. O. Andersen and T. Haugset, Phys. Rev. D **51**, 3073 (1995).
[33] A. S. Vshivtsev and K. G. Klimenko, Zh. Éksp. Teor. Fiz. **109**, 954 (1996) [JETP **82**, 514 (1996)]; A. S. Vshivtsev, K. G. Klimenko, and B. V. Magnitskiĭ, Zh. Éksp. Teor. Fiz. **107**, 307 (1995) [JETP **80**, 162 (1995)]; D. Ebert and A. S. Vshivtsev, Preprint HUB-EP-97/92, Berlin (1997).
[34] A. S. Vshivtsev and V. K. Peres-Fernandes, Dokl. Akad. Nauk SSSR **309**(1), 70 (1989) [Sov. Phys. Dokl. **34**, 968 (1989)].

Translated by Paul F. Schippnick

Tomography of two-particle spin states

V. A. Andreev and V. I. Man'ko*)

P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 10 March 1998)
Zh. Éksp. Teor. Fiz. **114**, 437–447 (August 1998)

A new ''classical'' formulation of conventional quantum mechanics is used to construct a description of quantum states with spin $j=1/2$ which uses only positive classical probabilities of observables instead of a complex density matrix. These probabilities are functions of points on a unit sphere which determine the directions relative to which the spin projections are measured. The proposed construction is a generalization of a system for tomography of the spin states of a single particle and may be used as a new method of measuring two-particle spin states. For pure spin states this method is exactly the same as the case of a single particle. Additional probabilities must be defined to describe mixed states. © *1998 American Institute of Physics.* [S1063-7761(98)00508-3]

1. INTRODUCTION

Since quantum mechanics was first developed and the Schrödinger equation[1] was formulated for the complex wave function of the coordinate, attempts have been made to apply a classical treatment to quantum-mechanical relationships.[2–5] A particular role which distinguishes the internal angular momentum from the coordinate is played by the particle spin introduced by Pauli,[6] an internal degree of freedom which has no classical analog. For pure states the spin state is described by complex spinors. The concept of the density matrix[7,8] was introduced to allow for quantum fluctuations in descriptions of the coordinates and the spins. The off-diagonal elements of the density matrix may be complex numbers, while the diagonal elements are always positive and have the meaning of the probabilities of the values of certain observables.

In Refs. 9 and 10, which were based on the results of Ref. 11, attention was drawn to the existence of a reversible transformation which links the Wigner function[12] of the coordinates and momenta with the positive distribution of probabilities defined by the diagonal elements of the same density matrix used to construct the Wigner function. In Ref. 13 this method was used to propose a new method of measuring the quantum state, known as optical tomography. It was shown in Refs. 14 and 15 that this circumstance may be used as the basis for a new quantum-mechanics formulation which uses a positive probability distribution rather than wave functions and density matrices. In fact, since two reversible transformations exist, one linking the density matrix with the Wigner function and the other linking the Wigner function with the probabilities of the measurable quantities, these probabilities contain the same information as the density matrix and may be used to describe the state of a quantum system. In other words, the density matrix can be reconstructed in terms of its diagonal elements measured in all reference frames.[14,15]

This approach was developed in Refs. 16–18 and a Fokker–Planck equation was obtained to describe the evolution of positive classical probabilities, and also its steady-state analog which can be used to find the energy levels of a system in terms of its ''classical'' distribution function. These serve as an alternative to the time-independent and time-dependent Schrödinger equations and were used in a new quantum-mechanics formulation to analyze an oscillator with friction.[19,20]

These equations describe systems of scalar particles for which the Wigner functions depend on the coordinates and the momenta. If the particles possess internal degrees of freedom such as spin, the formalism developed in Refs. 14 and 15 requires substantial modification. The problem of generalizing the probability formulation to the spin states of a single particles was examined in Refs. 21, 22, and 26. It was shown that the spin state may be defined by the probability distribution of the spin projection on a selected axis whose direction is fixed by two angles which determine the point on a unit sphere through which it passes. The direct generalization of this description to two particles with spin using two sets of independent probabilities which are functions of the coordinates of the points on two different spheres was considered in Refs. 23 and 24.

However, the problem of measuring the spin state of two or more particles with spin is best solved by using the same approach as that for a single particle. For a single particle the state can be completely measured if the probability of the spin projection on a selected axis can be determined experimentally in all reference frames. Thus, we need to be able to describe the state of two particles with spin such that the states of the individual spins are defined not by independent probabilities of the spin projections on different axes with independent rotations of the measurement frames for each spin, but by a general probability distribution which depends on a single rotation common to both spins.

This problem is solved here. We shall show that the pure state of two particles with spin $j=1/2$ can be completely defined by the distribution of the probabilities of the spin projections on their axes considered in an ensemble of systems turned with a single common rotation and we shall also

analyze the difference between the mixed and pure states encountered for many-particle spin systems. This difference arises from the fact that for many-particle spin systems, to obtain a complete description of the mixed spin state we need to know not only the probabilities of the spin projections as functions of the coordinates of the points on a unit sphere but also the probabilities defining the contributions of the pure states to the mixed ones. The mathematical basis of this method involves constructing a reversible transformation (''change of variables'') between the spin density matrix and the positive probabilities. This formulation is of interest because of the possibility of a new approach to the Einstein–Podolsky–Rosen paradox and hidden variables for spin degrees of freedom and also because of the possibility of a new ''tomographic'' method of measuring the spin state of many particles, among other applications.

2. TRANSFORMATIONS OF ELEMENTS OF THE TWO-PARTICLE DENSITY MATRIX

We shall analyze the density matrix of two-particle states with arbitrary spins and we shall show that its off-diagonal elements may be expressed in terms of the diagonal elements measured in all reference frames.

Let us assume that we have single-particle states with spins j_1 and j_2 characterized by the wave functions $\Psi^{j_1}_{m_1}$ and $\Psi^{j_2}_{m_2}$. These can be used to construct the two-particle state and its wave function is expressed in terms of the products $\Psi^{j_1}_{m_1}\Psi^{j_2}_{m_2}$. For pure two-particle states the elements of the density matrix are expressed in terms of the wave function:

$$\rho^{j_1 j_2}_{(m_1 m_2)(m_1' m_2')} = (\Psi^{j_1}_{m_1}\Psi^{j_2}_{m_2})(\Psi^{j_1}_{m_1'}\Psi^{j_2}_{m_2'})^*. \tag{1}$$

Assuming that the wave functions of the single-particle spin states are transformed in accordance with $O(3)$ rotation group representations, expression (1) may be transformed to give

$$\rho^{j_1 j_2}_{(m_1 m_2)(m_1' m_2')} = \sum_{l,k=|j_1-j_2|}^{j_1+j_2} C^{j_1\, j_2\, l}_{m_1 m_2 m}\; C^{j_1\, j_2\, k}_{m_1' m_2' m'}\; \Psi^{l}_{m}\, \Psi^{k\,*}_{m'}\quad , \tag{2}$$

where $C^{j_1\, j_2\, j}_{m_1 m_2 m}$ are the Clebsch–Gordan coefficients of the $O(3)$ group.

When we go to a different coordinate system, the product $\Psi^{l}_{m}\Psi^{k\,*}_{m'}$ is transformed as follows:

$$\tilde{\Psi}^{l}_{\tilde{m}}\tilde{\Psi}^{k\,*}_{\tilde{m}'} = \sum_{m=-l}^{l}\sum_{m'=-k}^{k} D^{l}_{\tilde{m}m}\Psi^{l}_{m}\;\Psi^{k*}_{m'}\, D^{k\,*}_{\tilde{m}'m'}\quad , \tag{3}$$

where D^{l}_{mn} is the Wigner D function.[25]

From this it follows that in a different coordinate system, the entire matrix element $\rho^{j_1, j_2}_{(m_1 m_2)(m_1' m_2')}$ is transformed according to the rule

$$\tilde{\rho}^{j_1 j_2}_{(\tilde{m}_1\tilde{m}_2)(\tilde{m}_1'\tilde{m}_2')} = \sum_{l,k} C^{j_1\, j_2\, l}_{\tilde{m}_1\tilde{m}_2\tilde{m}}\; C^{j_1\, j_2\, k}_{\tilde{m}_1\tilde{m}_2\tilde{m}}\sum_{m,m'} D^{l}_{\tilde{m}m}\rho^{l\,k}_{mm'}D^{k\,*}_{\tilde{m}'m'}\quad , \tag{4}$$

where $\rho^{l\,k}_{mm'} = \Psi^{l}_{m}\Psi^{k\,*}_{m'}$. This transformation rule (4) is valid in the general case when the state is not pure and the density matrix elements do not have the form (1). In this case, the values of $\rho^{l\,k}_{mm'}$ are determined using the relation

$$\rho^{l\,k}_{mm'} = \sum_{m_1, m_2, m_1', m_2'} C^{j_1\, j_2\, l}_{m_1 m_2 m}\; C^{j_1\, j_2\, k}_{m_1' m_2' m'}\; \rho^{j_1\, j_2}_{(m_1 m_2)(m_1' m_2')}\quad . \tag{5}$$

Knowing the values of $\rho^{l\,k}_{mm'}$, we can reconstruct the matrix elements $\rho^{j_1 j_2}_{(m_1 m_2)(m_1' m_2')}$:

$$\rho^{j_1 j_2}_{(m_1 m_2)(m_1' m_2')} = \sum_{l,k} C^{j_1\, j_2\, l}_{m_1 m_2 m}\; C^{j_1\, j_2\, k}_{m_1' m_2' m'}\; \rho^{l\,k}_{mm'}\quad . \tag{6}$$

When we go to a different coordinate system, the values of $\rho^{l\,k}_{mm'}$ transform as

$$\tilde{\rho}^{l\,k}_{\tilde{m}\tilde{m}'} = \sum_{m,m'} D^{l}_{\tilde{m}m}(\phi,\theta,\psi)\;\rho^{l\,k}_{mm'}\, D^{k\,*}_{\tilde{m}'m'}(\phi,\theta,\psi), \tag{7}$$

and the matrix elements $\rho^{j_1 j_2}_{(m_1 m_2)(m_1' m_2')}$ transform in accordance with formula (4).

Our aim is to derive a formula which expresses the off-diagonal elements of the density matrix in terms of the diagonal ones measured in all coordinate systems. It is difficult to relate these matrix elements directly, and thus we shall first express the values of $\rho^{l\,k}_{mm'}$ in terms of the diagonal matrix elements and then, using formula (6), we shall reconstruct all $\rho^{j_1 j_2}_{(m_1 m_2)(m_1' m_2')}$ using these. For this purpose we take the diagonal element $\rho^{j_1\, j_2}_{(\tilde{m}_1\tilde{m}_2)(\tilde{m}_1\tilde{m}_2)}(\phi,\theta,\psi)$ and express it in terms of $\rho^{l\,k}_{mm'}$, relative to some fixed coordinate system:

$$\begin{aligned}
\tilde{\rho}^{j_1\, j_2}_{(\tilde{m}_1\tilde{m}_2)(\tilde{m}_1\tilde{m}_2)} &= \sum_{l,k} C^{j_1\, j_2\, l}_{\tilde{m}_1\tilde{m}_2\tilde{m}}\; C^{j_1\, j_2\, k}_{\tilde{m}_1\tilde{m}_2\tilde{m}}\,\tilde{\rho}^{l\,k}_{\tilde{m}\tilde{m}} \\
&= \sum_{l,k} C^{j_1\, j_2\, l}_{\tilde{m}_1\tilde{m}_2\tilde{m}}\; C^{j_1\, j_2\, k}_{\tilde{m}_1\tilde{m}_2\tilde{m}} \\
&\quad\times \sum_{m,m'} D^{l}_{\tilde{m}'m}(\phi,\theta,\psi)\;\rho^{l\,k}_{mm'}\; D^{k\,*}_{\tilde{m}'m'}(\phi,\theta,\psi) \\
&= \sum_{l,k} C^{j_1\, j_2\, l}_{\tilde{m}_1\tilde{m}_2\tilde{m}}\; C^{j_1\, j_2\, k}_{\tilde{m}_1\tilde{m}_2\tilde{m}} \\
&\quad\times \sum_{m,m',S} C^{l\;k\;S}_{\tilde{m}\;-\tilde{m}\;0}\; C^{l\;k\;S}_{m\;-m'\;m-m'}(-1)^{\tilde{m}-m'} \\
&\quad\times D^{S}_{0,m-m'}(\phi,\theta,\psi)\;\rho^{l\;k}_{m\;m'}.
\end{aligned} \tag{8}$$

We multiply the right- and left-hand sides of Eq. (8) by $D^{S}_{0,m-m'}$ and integrate over all angles:

$$\frac{1}{\pi^2}\int d\Omega\ \tilde{\rho}^{j_1\,j_2}_{(\tilde{m}_1\tilde{m}_2)(\tilde{m}_1\tilde{m}_2)} D^S_{0,m-m'}(\phi,\theta,\psi)(2S+1)(-1)^{-\tilde{m}}$$

$$=\sum_{l,k} C^{j_1\,j_2\,l}_{\tilde{m}_1\tilde{m}_2\tilde{m}}\ C^{j_1\,j_2\,k}_{\tilde{m}_1\tilde{m}_2\tilde{m}} \sum_{m,m'}\ C^{l\ \ k\ \ S}_{\tilde{m}\ -\tilde{m}\ 0}\, C^{l\ \ k\ \ S}_{m\ -m'\ m-m'}$$

$$\times(-1)^{-m'}\rho^{l\ k}_{m\ m'}=P^{j_1,\,j_2,\ S}_{\tilde{m}_1,\tilde{m}_2,m-m'}\quad . \tag{9}$$

In order to reconstruct all the elements of the density matrix using formula (6), we must solve the system (9) and express $\rho^{l\,k}_{m\,m'}$ in terms of the diagonal elements $\rho^{j_1\,j_2}_{(\tilde{m}_1\tilde{m}_2)(\tilde{m}_1\tilde{m}_2)}$. However, it seems that unlike single-particle states, the density matrix of many-particle states is not always completely reconstructed in terms of its diagonal elements. We shall analyze this problem in detail for the case $j_1=j_2=1/2$, and in the general case we shall calculate those elements $\rho^{l\,k}_{mm'}$ for which an explicit expression can be derived without any additional assumptions, these being the values $\rho^{l\,l}_{m\,m'}$. In order to calculate these, we sum the system (9) over $\tilde{m}_1$ and $\tilde{m}_2$, keeping $\tilde{m}$ constant. Taking into account that

$$\sum_{m_1,m_2} C^{j_1\,j_2\,l}_{m_1m_2m}\ C^{j_1\,j_2\,k}_{m_1m_2m}=\delta_{lk}\quad , \tag{10}$$

we obtain

$$\sum_{\tilde{m}_1,\tilde{m}_2} P^{j_1\,j_2\,S}_{\tilde{m}_1\tilde{m}_2m-m'}=\sum_l\ \sum_{m,m'}\ (-1)^{-m'}$$

$$\times C^{l\ \ l\ \ S}_{\tilde{m}-\tilde{m}\ 0}\ \ C^{l\ \ l\ \ S}_{m\ -m'\ m-m'}\ \ \rho^{l\,l}_{m\,m'}$$

$$=Q^{j_1,\,j_2,\,S}_{\tilde{m},m-m'}\quad . \tag{11}$$

If we assume that $\tilde{m}=\tilde{m}_1+\tilde{m}_2=j_1+j_2,$, then the sum over l in Eq. (11) only contains one term

$$Q^{j_1,j_2,S}_{j_1+j_2,m-m'}=C^{j_1+j_2\ j_1+j_2\ S}_{j_1+j_2\ -j_1-j_2\ 0}$$

$$\times\sum_{m,m'} C^{j_1+j_2\ j_1+j_2\ \ S}_{m\ \ \ -m'\ \ m-m'}$$

$$\times(-1)^{-m'}\rho^{j_1+j_2\ j_1+j_2}_{m\ m'}\quad . \tag{12}$$

Here $m-m'$ is fixed. Using the identity

$$\sum_S\ C^{j_1j_2\ \ S}_{m_1\,m_2\,m}\ \ C^{j_1\ \ \ j_2\ S}_{m_1'\,m_2'\,m'}=\delta_{m_1m_1'}\ \delta_{m_2m_2'}\quad , \tag{13}$$

we obtain from Eq. (12)

$$\rho^{j_1+j_2\ j_1+j_2}_{m\ m'}=\sum_S\ (-1)^{m'}Q^{j_1,j_2,S}_{j_1+j_2,m-m'}$$

$$\times(C^{j_1+j_2\ j_1+j_2\ S}_{j_1+j_2\ -j_1-j_2\ 0})^{-1}C^{j_1+j_2\ j_1+j_2\ \ S}_{m\ \ \ -m'\ m-m'}\quad . \tag{14}$$

All the other values $\rho^{l\,l}_{m\,m'}$ can be calculated similarly; in the process the parameter m must be reduced step by step in Eq. (11). The resulting equation will contain one unknown value of $\rho^{l\,l}_{m\,m'}$ and several known values. The orthogonality relation (13) can be used to express the unknown in terms of the known values. To obtain the required formula, we introduce the notation

$$\Phi^S_n=\sum_{l=n+1}^{j_1+j_2} C^{l\ \ l\ S}_{n\ -n\ 0}\sum_{m,m'}\ C^{l\ \ l\ \ S}_{m\ -m'\ m-m'}(-1)^{m'}\rho^{l\,l}_{m\,m'}\quad . \tag{15}$$

Using this, we obtain

$$\rho^{n\,n}_{m\,m'}=\sum_S\ (Q^{j_1,\,j_2,\,S}_{n,\,m-m'}-\Phi^S_n)$$

$$\times(C^{n\ n\ S}_{n\ -n\ 0})^{-1}C^{n\ \ n\ \ S}_{m\ \ -m'\ \ m-m'}(-1)^{m'}\quad . \tag{16}$$

3. TOMOGRAPHY OF PARTICLES WITH SPIN $j=1/2$

We shall now analyze the case $j_1=j_2=1/2$. Using formula (16), we can calculate ten values

$$\rho^{00}_{00}\quad ,\quad \rho^{11}_{mm'},\quad m,m'=-1,0,1 \tag{17}$$

leaving six values of ρ^{10}_{m0}, ρ^{01}_{0m} still to be determined.

However, it is easy to establish that these appear as linear combinations in formula (9). For example, the right-hand side $P^{1/2\,1/2\,1}_{1/2\,-1/2\,0}$ contains the sum $\rho^{10}_{00}+\rho^{01}_{00}$, the right-hand side $P^{1/2\,1/2\,1}_{1/2\,-1/2\,1}$ contains the difference $\rho^{1\,0}_{1\,0}-\rho^{0\,1}_{0\,-1}$, and the right-hand side $P^{1/2\,1/2\,1}_{1/2\,-1/2\,-1}$ contains the difference $\rho^{1\,0}_{-1\,0}-\rho^{0\,1}_{0\,1}$. In the formulas (9) with the values $P^{1/2\,1/2\,1}_{-1/2\,1/2\,0}$, $P^{1/2\,1/2\,1}_{-1/2\,1/2\,1}$, and $P^{1/2\,1/21}_{-1/2\,1/2\,-1}$, the same linear combinations appear but with opposite sign

$$(-\rho^{10}_{00}-\rho^{01}_{00}),(-\rho^{10}_{10}+\rho^{0\,1}_{0-1}),(-\rho^{10}_{-10}+\rho^{01}_{10}). \tag{18}$$

Thus, ten values of (17) and three values of (18) are reconstructed in terms of the diagonal elements of the density matrix. This is insufficient to reconstruct the entire density matrix.

We can immediately check that if we take the off-diagonal matrix elements $\rho^{1/2\,1/2}_{(1/2-1/2)(-1/2\,1/2)}$ and $\rho^{1/2\,1/2}_{(-1/21/2)(1/2-1/2)}$ and write a formula similar to (9) for these, the following linear combinations appear on the right-hand side of the resulting relations

$$(\rho^{10}_{00}-\rho^{01}_{00}),\quad (\rho^{10}_{-10}+\rho^{01}_{01}),\quad (\rho^{10}_{10}+\rho^{0\,1}_{0-1}),$$

$$-(\rho^{10}_{00}-\rho^{01}_{00}),-(\rho^{10}_{-10}+\rho^{01}_{01}),\ \ -(\rho^{10}_{10}+\rho^{0\,1}_{0-1}). \tag{19}$$

We can see that having only some of the diagonal elements of the density matrix is insufficient to reconstruct the entire density matrix; further information on the state is required. Let us assume, for example, that the state is pure. The elements of its density matrix then satisfy

$$\rho_{nn}\rho_{mm}=\rho_{nm}\rho_{mn}\,. \tag{20}$$

We substitute into Eq. (20) the expressions for the matrix elements

$$\rho^{1/2\,1/2}_{(1/2\,1/2)(1/2\,1/2)},\quad \rho^{1/2\,1/2}_{(1/2\,-1/2)(1/2\,-1/2)},\quad \rho^{1/2\,1/2}_{(1/2\,-1/2)(1/2\,1/2)},$$

$$\rho^{1/2\,1/2}_{(1/2\,1/2)(1/2\,-1/2)},\quad \rho^{1/2\,1/2}_{(-1/2\,1/2)(-1/2\,1/2)},\quad \rho^{1/2\,1/2}_{(1/2\,-1/2)(-1/2\,1/2)},$$

$$\rho^{1/2\,1/2}_{(-1/2\,1/2)(1/2\,-1/2)},\quad \rho^{1/2\,1/2}_{(1/2\,1/2)(-1/2\,1/2)},\quad \rho^{1/2\,1/2}_{(-1/2\,1/2)(1/2\,1/2)} \tag{21}$$

in terms of the values $\rho^{l\,k}_{m\,n}$, using the formulas from Appendix 1, and we obtain

$$(\rho^{1\,0}_{0\,0}-\rho^{0\,1}_{0\,0})^2=(\rho^{1\,0}_{0\,0}+\rho^{0\,1}_{0\,0})^2-4\rho^{0\,0}_{0\,0}\rho^{1\,1}_{0\,0}. \tag{22}$$

Since $\rho^{10}_{00}+\rho^{01}_{00}$ is determined by the diagonal terms of the density matrix, Eq. (22) can be used to calculate ρ^{10}_{00} and ρ^{01}_{00}. In addition to Eq. (22), the relations (20) yield

$$\rho^{11}_{11}(\rho^{00}_{00}+\rho^{11}_{00}+\rho^{10}_{00}+\rho^{01}_{00})=(\rho^{11}_{10}+\rho^{10}_{10})(\rho^{11}_{01}+\rho^{01}_{01}), \tag{23}$$

$$\rho^{11}_{11}(\rho^{00}_{00}+\rho^{11}_{00}-\rho^{10}_{00}-\rho^{01}_{00})=(\rho^{11}_{10}-\rho^{10}_{10})(\rho^{11}_{01}-\rho^{01}_{01}). \tag{24}$$

By adding and subtracting Eqs. (23) and (24), we obtain

$$\rho^{10}_{10}\rho^{01}_{01}=\rho^{11}_{11}(\rho^{00}_{00}+\rho^{11}_{00})-\rho^{11}_{10}\rho^{11}_{01}, \tag{25}$$

$$\rho^{10}_{10}\rho^{11}_{01}+\rho^{01}_{01}\rho^{11}_{10}=\rho^{11}_{11}(\rho^{10}_{00}+\rho^{01}_{00}). \tag{26}$$

Solving the systems (25) and (26), we can find ρ^{10}_{10} and ρ^{01}_{01}, which immediately gives us $\rho^{0\,1}_{0\,-1}$ and ρ^{10}_{-10}.

Thus, the relations (20) obtained for pure states can be used to reconstruct the entire density matrix of the pure state in terms of its diagonals.

We shall now analyze a mixed state formed by the pure states

$$\begin{aligned}&\Psi_1=\alpha|(1)+\rangle|(2)-\rangle+\beta|(1)-\rangle|(2)+\rangle,\\ &\Psi_2=\beta^*|(1)+\rangle|(2)-\rangle-\alpha^*|(1)-\rangle|(2)+\rangle,\\ &\alpha\alpha^*+\beta\beta^*=1,\end{aligned} \tag{27}$$

contained in it with the probabilities w_1 and w_2.

The density matrix of the state Ψ_1 has the form

$$\rho(\Psi_1)=\begin{pmatrix}0&0&0&0\\0&\alpha\alpha^*&\alpha\beta^*&0\\0&\beta\alpha^*&\beta\beta^*&0\\0&0&0&0\end{pmatrix}. \tag{28}$$

In a rotated system of coordinates the elements of the density matrix (28) are expressed by the following formulas:

$$\rho_{11}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(\frac12\,\frac12\right)\left(\frac12\,\frac12\right)}=\frac14(\alpha+\beta)(\alpha^*+\beta^*)\sin^2\theta,$$

$$\rho_{12}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(\frac12\,\frac12\right)\left(\frac12-\frac12\right)}=\frac{i}{2}e^{-i\varphi}(\alpha+\beta)\times\sin\theta\left(\alpha^*\cos^2\frac{\theta}{2}-\beta^*\sin^2\frac{\theta}{2}\right),$$

$$\rho_{13}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(\frac12\,\frac12\right)\left(-\frac12\,\frac12\right)}=\frac{i}{2}e^{-i\varphi}(\alpha+\beta)\times\sin\theta\left(\beta^*\cos^2\frac{\theta}{2}-\alpha^*\sin^2\frac{\theta}{2}\right),$$

$$\rho_{14}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(\frac12\,\frac12\right)\left(-\frac12-\frac12\right)}=\frac14e^{-2i\varphi}(\alpha+\beta)\times(\alpha^*+\beta^*)\sin^2\theta,$$

$$\rho_{22}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(\frac12-\frac12\right)\left(\frac12-\frac12\right)}=\left(\alpha\cos^2\frac{\theta}{2}-\beta\sin^2\frac{\theta}{2}\right)\times\left(\alpha^*\cos^2\frac{\theta}{2}-\beta^*\sin^2\frac{\theta}{2}\right),$$

$$\rho_{23}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(\frac12-\frac12\right)\left(-\frac12\,\frac12\right)}=\left(\alpha\cos^2\frac{\theta}{2}-\beta\sin^2\frac{\theta}{2}\right)\times\left(\beta^*\cos^2\frac{\theta}{2}-\alpha^*\sin^2\frac{\theta}{2}\right),$$

$$\rho_{24}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(\frac12-\frac12\right)\left(-\frac12-\frac12\right)}=-\frac{i}{2}e^{-i\varphi}\times(\alpha^*+\beta^*)\sin\theta\left(\alpha\cos^2\frac{\theta}{2}-\beta\sin^2\frac{\theta}{2}\right),$$

$$\rho_{34}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(-\frac12\,\frac12\right)\left(-\frac12-\frac12\right)}=\frac{i}{2}e^{-i\varphi}\times(\alpha^*+\beta^*)\sin\theta\left(\alpha\sin^2\frac{\theta}{2}-\beta\cos^2\frac{\theta}{2}\right),$$

$$\rho_{44}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(-\frac12-\frac12\right)\left(-\frac12-\frac12\right)}=\frac14(\alpha+\beta)\times(\alpha^*+\beta^*)\sin^2\theta,$$

$$\rho_{33}(\Psi_1)=\rho^{\frac12\,\frac12}_{\left(-\frac12\,\frac12\right)\left(-\frac12\,\frac12\right)}=\left(\alpha\sin^2\frac{\theta}{2}-\beta\cos^2\frac{\theta}{2}\right)\times\left(\alpha^*\sin^2\frac{\theta}{2}-\beta^*\cos^2\frac{\theta}{2}\right). \tag{29}$$

The elements of the density matrix of the state Ψ_2 are similar.

The elements of the density matrix of the mixed state are the sums of the density matrix elements of its constituent states, multiplied by the appropriate probabilities.

We introduce the notation

$$\alpha=\sin\omega e^{i\tau},\quad \beta=\cos\omega e^{-i\tau}. \tag{30}$$

Using the parameterization (30), we write the diagonal elements of the density matrix of the mixed state formed by the pure states (27)

$$\rho_{11}=\frac14(1+\Delta\sin2\omega\cos2\tau)\sin^2\theta,$$

$$\rho_{22}=\frac12\left(1-\Delta\cos2\omega\cos\theta-\frac12\times\sin^2\theta(1+\Delta\sin2\omega\cos2\tau)\right),$$

$$\rho_{33}=\frac12\left(1+\Delta\cos2\omega\cos\theta-\frac12\times\sin^2\theta(1+\Delta\sin2\omega\cos2\tau)\right),$$

$$\rho_{44}=\frac{1}{4}(1+\Delta\sin 2\omega\cos 2\tau)\sin^2\theta, \tag{31}$$

where $\Delta=w_1-w_2$.

We can see that the diagonal matrix elements contain two quantities $Q_1=\Delta\sin2\omega\cos2\tau$ and $Q_2=\Delta\cos2\omega$ which depend on three parameters. The elements ρ_{14} and ρ_{41} are expressed in terms of these two parameters, and all the others depend on the value of $R=\Delta\sin2\omega\sin2\tau$, which is not expressed in terms of Q_1 and Q_2. We can easily derive the expressions

$$R=(\Delta^2-Q_1^2-Q_2^2)^{1/2},\quad \cos2\omega=Q_2/\Delta,$$
$$\cos2\tau=Q_1(\Delta^2-Q_2^2)^{-1/2}. \tag{32}$$

The formulas (32) describe a single-parameter family of states (27) whose density matrices have the same diagonal elements (31) but different off-diagonal ones. In general, when a mixed state consists of four pure ones, the states having the same diagonal density matrix elements belong to a three-parameter family. This is easily confirmed by calculating the number of parameters which are not determined using formula (9). We cannot determine the values (19), these being three complex numbers, but from formula (5) we can easily obtain

$$(\rho_{00}^{01})^*=\rho_{00}^{01},\quad (\rho_{01}^{01})^*=-\rho_{-10}^{10},\quad (\rho_{0-1}^{01})^*=\rho_{10}^{10}\ . \tag{33}$$

It follows from the formulas (33) that the unknowns (19) contain only three real unknown parameters and these three parameters determine the family of states. In order to identify a specific state inside a family, we need to define three real numbers characterizing these, for which we can take the probabilities w_1, w_2, w_3, and w_4 which define the structure of the mixed state.

Defining the probabilities $w_1=1$, $w_2=w_3=w_4=0$, we obtain the pure state and for a mixed state consisting of pure states, we have $w_1\neq0$, $w_2\neq0$, $w_3=0$, $w_4=0$, $w_1+w_2=1$ and we arrive at the formulas (32) which can reconstruct the entire density matrix. Similar but more complex formulas can be obtained for the case when all $w_i\neq0$, $i=1,\ldots,4$.

4. CONCLUSIONS

A concept whereby the entire quantum mechanics formalism was reformulated using only measurable quantities was developed in Refs. 14–18. It was suggested that the classical probabilities of the values of particular observables should be used as these quantities. For the systems analyzed in Refs. 14–18 these probabilities are defined by the diagonal elements of the density matrix of the corresponding states.

The results of the present study show that for mixed two-particle states, using only some of the diagonal elements of the density matrix examined merely in terms of a single reference frame common to both spins is inadequate for a complete description of the state and these must be supplemented by the probabilities $w_i, i=1\ldots$, with which the pure states appear in the mixed ones. The diagonal elements are then used to construct a family of states whose density matrices have the same diagonal matrix elements but different off-diagonal ones and the probabilities w_i determine the required state within the family. Thus, we have implemented the overall concept of defining states using classical probabilities.

This work was supported financially by the Russian Fund for Fundamental Research (Grants Nos. 96-02-17222 and 96-02-17987).

APPENDIX A:

We give the expressions for the elements of the two-particle density matrix for spins $j_1=j_2=1/2$ in terms of ρ_{mn}^{kl}:

$$\rho^{\frac12\frac12}_{\left(\frac12\frac12\right)\left(\frac12\frac12\right)}=\rho_{11}^{11},$$

$$\rho^{\frac12\frac12}_{\left(\frac12\frac12\right)\left(\frac12-\frac12\right)}=\frac{1}{\sqrt2}(\rho_{10}^{11}+\rho_{10}^{10}),$$

$$\rho^{\frac12\frac12}_{\left(\frac12\frac12\right)\left(-\frac12\frac12\right)}=\frac{1}{\sqrt2}(\rho_{10}^{11}-\rho_{10}^{10}),$$

$$\rho^{\frac12\frac12}_{\left(\frac12\frac12\right)\left(-\frac12-\frac12\right)}=\rho_{1-1}^{11},$$

$$\rho^{\frac12\frac12}_{\left(\frac12-\frac12\right)\left(\frac12\frac12\right)}=\frac{1}{\sqrt2}(\rho_{01}^{11}-\rho_{01}^{01}),$$

$$\rho^{\frac12\frac12}_{\left(\frac12-\frac12\right)\left(\frac12-\frac12\right)}=\frac12(\rho_{00}^{00}+\rho_{00}^{11}+\rho_{00}^{10}+\rho_{00}^{01}),$$

$$\rho^{\frac12\frac12}_{\left(\frac12-\frac12\right)\left(-\frac12\frac12\right)}=\frac12(-\rho_{00}^{00}+\rho_{00}^{11}-\rho_{00}^{10}+\rho_{00}^{01}),$$

$$\rho^{\frac12\frac12}_{\left(\frac12-\frac12\right)\left(-\frac12-\frac12\right)}=\frac{1}{\sqrt2}(\rho_{0-1}^{11}+\rho_{0-1}^{01}),$$

$$\rho^{\frac12\frac12}_{\left(-\frac12\frac12\right)\left(\frac12\frac12\right)}=\frac{1}{\sqrt2}(\rho_{01}^{11}-\rho_{01}^{01}),$$

$$\rho^{\frac12\frac12}_{\left(-\frac12\frac12\right)\left(\frac12-\frac12\right)}=\frac12(-\rho_{00}^{00}+\rho_{00}^{11}+\rho_{00}^{10}-\rho_{00}^{01}),$$

$$\rho^{\frac12\frac12}_{\left(-\frac12\frac12\right)\left(-\frac12\frac12\right)}=\frac12(\rho_{00}^{00}+\rho_{00}^{11}-\rho_{00}^{10}-\rho_{00}^{01}),$$

$$\rho^{\frac12\frac12}_{\left(-\frac12\frac12\right)\left(-\frac12-\frac12\right)}=\frac{1}{\sqrt2}(\rho_{0-1}^{11}-\rho_{0-1}^{01}),$$

$$\rho^{\frac12\frac12}_{\left(-\frac12-\frac12\right)\left(\frac12\frac12\right)}=\rho_{-11}^{11},$$

$$\rho^{\frac12\frac12}_{\left(-\frac12-\frac12\right)\left(\frac12-\frac12\right)}=\frac{1}{\sqrt2}(\rho_{-10}^{11}-\rho_{-10}^{10}),$$

$$\rho^{\frac{1}{2}\frac{1}{2}}_{\left(-\frac{1}{2}-\frac{1}{2}\right)\left(-\frac{1}{2}\frac{1}{2}\right)}=\frac{1}{\sqrt{2}}(\rho^{11}_{-10}+\rho^{10}_{-10}),$$

$$\rho^{\frac{1}{2}\frac{1}{2}}_{\left(-\frac{1}{2}-\frac{1}{2}\right)\left(-\frac{1}{2}-\frac{1}{2}\right)}=\rho^{11}_{-1-1}.$$

APPENDIX B:

We give the explicit form of the elements of the density matrix of various two-particle states. These are calculated using formula (3) and the D-functions

$$D^{\frac{1}{2}}_{\frac{1}{2}\frac{1}{2}}=e^{-i\frac{1}{2}(\phi+\psi)}\cos\frac{\theta}{2},\quad D^{\frac{1}{2}}_{\frac{1}{2}-\frac{1}{2}}=ie^{-i\frac{1}{2}(\phi-\psi)}\sin\frac{\theta}{2},$$

$$D^{\frac{1}{2}}_{-\frac{1}{2}\frac{1}{2}}=ie^{-i\frac{1}{2}(-\phi+\psi)}\cos\frac{\theta}{2},$$

$$D^{\frac{1}{2}}_{-\frac{1}{2}-\frac{1}{2}}=ie^{-i\frac{1}{2}(-\phi-\psi)}\sin\frac{\theta}{2}.$$

The singlet state is given by:

$$\Psi_{00}=\frac{1}{\sqrt{2}}(|(1)+\rangle|(2)-\rangle-|(1)-\rangle|(2)+\rangle),$$

$$\rho_{22}=\frac{1}{2},\quad \rho_{23}=-\frac{1}{2},\quad \rho_{32}=-\frac{1}{2},\quad \rho_{33}=\frac{1}{2}.$$

The other matrix elements are zero. Since the state Ψ_{00} is a scalar, the elements of its density matrix remain unchanged when we go to a different coordinate system, as is confirmed by direct calculations using formula (3).

The triplet state is given by:

$$\Psi_{10}=\frac{1}{\sqrt{2}}(|(1)+\rangle|(2)-\rangle+|(1)-\rangle|(2)+\rangle),$$

$$\Psi_{11}=|(1)+\rangle|(2)+\rangle,\quad \Psi_{1-1}=|(1)-\rangle|(2)-\rangle.$$

In the laboratory coordinate system the elements of the density matrix of the state Ψ_{10} have the form

$$\rho_{22}=\frac{1}{2},\quad \rho_{23}=\frac{1}{2},\quad \rho_{32}=\frac{1}{2},\quad \rho_{33}=\frac{1}{2},$$

and the others are zero. On changing to a different coordinate system in accordance with formula (3), we have

$$\tilde{\rho}_{11}=\frac{1}{2}\sin^2\theta,\quad \tilde{\rho}_{12}=\tilde{\rho}_{13}=\frac{i}{2}e^{-i\phi}\sin\theta\cos\theta,$$

$$\tilde{\rho}_{14}=\frac{1}{2}e^{-2i\phi}\sin^2\theta,$$

$$\tilde{\rho}_{22}=\tilde{\rho}_{23}=\frac{1}{2}\cos^2\theta,\quad \tilde{\rho}_{24}=-\frac{i}{2}e^{-i\phi}\sin\theta\cos\theta,$$

$$\tilde{\rho}_{33}=\frac{1}{2}\cos^2\theta,\quad \tilde{\rho}_{34}=-\frac{i}{2}e^{-i\phi}\sin\theta\cos\theta,$$

$$\tilde{\rho}_{44}=\frac{1}{2}\sin^2\theta.$$

The state Ψ_{11} corresponds to a density matrix with a single nonzero element $\rho_{11}=1$. In an arbitrary coordinate system we have

$$\tilde{\rho}_{11}=\cos^4\frac{\theta}{2},\quad \tilde{\rho}_{12}=\tilde{\rho}_{13}=-ie^{-i\phi}\sin\frac{\theta}{2}\cos^3\frac{\theta}{2},$$

$$\tilde{\rho}_{14}=-ie^{-i\phi}\sin^2\theta\cos^2\theta,$$

$$\tilde{\rho}_{22}=\tilde{\rho}_{23}=\sin^2\frac{\theta}{2}\cos^2\frac{\theta}{2},\quad \tilde{\rho}_{24}=-ie^{-i\phi}\sin^3\frac{\theta}{2}\cos\frac{\theta}{2},$$

$$\tilde{\rho}_{33}=\sin^2\frac{\theta}{2}\cos^2\frac{\theta}{2},\quad \tilde{\rho}_{34}=-ie^{-i\phi}\sin^3\frac{\theta}{2}\cos\frac{\theta}{2},$$

$$\tilde{\rho}_{44}=\sin^4\frac{\theta}{2}.$$

For the state Ψ_{1-1} there is also only one nonzero matrix element $\rho_{44}=1$. When we go to an arbitrary coordinate system, we obtain

$$\tilde{\rho}_{11}=\sin^4\frac{\theta}{2},\quad \tilde{\rho}_{12}=\tilde{\rho}_{13}=ie^{-i\phi}\cos\frac{\theta}{2}\sin^3\frac{\theta}{2},$$

$$\tilde{\rho}_{14}=-ie^{-i\phi}\sin^2\theta\cos^2\theta,$$

$$\tilde{\rho}_{22}=\tilde{\rho}_{23}=\sin^2\frac{\theta}{2}\cos^2\frac{\theta}{2},\quad \tilde{\rho}_{24}=ie^{-i\phi}\cos^3\frac{\theta}{2}\sin\frac{\theta}{2},$$

$$\tilde{\rho}_{33}=\sin^2\frac{\theta}{2}\cos^2\frac{\theta}{2},\quad \tilde{\rho}_{34}=ie^{-i\phi}\cos^3\frac{\theta}{2}\sin\frac{\theta}{2},$$

$$\tilde{\rho}_{44}=\cos^4\frac{\theta}{2}.$$

*)E-mail: manko@astrna.na.astro.it

[1] E. Schrödinger, Ann. Phys. (Leipzig) **79**, 489 (1926).
[2] L. De Broglie, Compt. Rend. **183**, 447 (1926); **184**, 273 (1927); **185**, 380 (1927).
[3] E. Madelung, Z. Phys. **40**, 332 (1926).
[4] D. Bohm, Phys. Rev. **85**, 166, 180 (1952).
[5] E. Nelson, Phys. Rev. **150**, 1079 (1966).
[6] W. Pauli, Z. Phys. **31**, 765 (1925).
[7] L. D. Landau, Z. Phys. **45**, 430 (1927).
[8] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, transl. from the German (Princeton University Press, Princeton, N.J., 1955).
[9] J. Bertrand and P. Bertrand, Found. Phys. **17**, 397 (1987).
[10] K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).
[11] K. E. Cahill and R. J. Glauber, Phys. Rev. **177**, 1882 (1969).
[12] E. Wigner, Phys. Rev. **40**, 749 (1932).
[13] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).
[14] S. Mancini, V. I. Man'ko, and P. Tombesi, Phys. Lett. A **213**, 1 (1996).
[15] S. Mancini, V. I. Man'ko, and P. Tombesi, Found. Phys. **27**, 801 (1997).
[16] V. I. Man'ko, J. Russ. Laser Res. **17**, 579 (1996).
[17] O. V. Man'ko and V. I. Man'ko, J. Russ. Laser Res. **18**, 407 (1997).
[18] V. I. Man'ko, in *Symmetries in Science IX*, edited by B. Gruber and M. Ramer (Plenum Press, New York, 1997), p. 215.
[19] V. I. Man'ko and S. S. Safonov, J. Russ. Laser Res. **18**, 501 (1997).
[20] V. I. Man'ko and S. S. Safonov, Teor. Mat. Fiz. **112**, 467 (1997).
[21] V. V. Dodonov and V. I. Man'ko, Phys. Lett. A **229**, 335 (1997).

[22] O. V. Man'ko and V. I. Man'ko, Zh. Éksp. Teor. Fiz. **112**, 796 (1997) [JETP **85**, 430 (1997)].
[23] V. I. Man'ko and S. S. Safonov, Yad. Fiz.**61**, 658 (1998) [Phys. At. Nucl. **61**, 585 (1998)].
[24] V. I. Man'ko and S. S. Safonov, Teor. Mat. Fiz. **115**, 185 (1998).
[25] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Relativistic Quantum Theory* (Pergamon Press, Oxford, 1971) [Russian original, later ed. Parts 1 and 2, Nauka, Moscow 1974].
[26] G. S. Agarwal, Phys. Rev. A **57**, 671 (1998).

Translated by R. M. Durham

Exact quantum mechanical description of the motion of spin-1/2 particles and of spin motion in a uniform magnetic field

A. Ya. Silenko*)

Nuclear Problems Research Institute at Belorussian State University, 220080 Minsk, Belarus
(Submitted 13 August 1997)
Zh. Éksp. Teor. Fiz. **114**, 448–457 (August 1998)

The Dirac–Pauli equation is used to obtain the exact equation of spin motion for spin-1/2 particles with an anomalous magnetic moment in a constant and uniform magnetic field. Exact formulas are established for the angular velocity of the revolution of such particles along circular orbits and the rotation of the particle spin with respect to momentum. Finally, a quantum mechanical equation for the motion of the particles in a strong magnetic field is derived.
 [S1063-7761(98)00608-8]

1. INTRODUCTION

Only a limited number of exact solutions of the Dirac and Dirac–Pauli equations[1] that allow for the presence of an anomalous magnetic moment are known. A full coverage of these solutions can be found in the monograph by Bagrov *et al.*[2] Almost all exact solutions amounted to finding the energy levels and wave particles of spin-$\frac{1}{2}$ particles in an external field. Such problems were solved in the Dirac representation and in the Foldy–Wouthuysen (FW) representation.[3]

A distinctive feature of the FW representation is the quasidiagonal form of the Hamiltonian operator, which separates the equations for the two spinors. Hence one spinor is sufficient for completely describing the state of particles in the FW representation. What is important is that in the FW representation the operators are even, so that there is no need to separate the even part of operators. The special role of the FW representation also consists in the fact that we can associate the corresponding classical quantities with the operators given in this representation (say, the position operator $\mathbf{r}$ and momentum operator $\mathbf{p}$), while as a result of unitary transformations to other representations, including the Dirac representation, the form of the given operators may change substantially.[3]

However, deriving the exact expressions for the Hamiltonian operator in the FW representation is extremely difficult, with the result that such expressions have been found only in some cases.[4–7] Among the problems that can be exactly solved in the FW representation is that of a particle with an anomalous magnetic moment μ' in a uniform magnetic field of arbitrary strength.[6,7] The exact equation of spin motion for the particular case of a Dirac particle ($\mu'=0$) in a uniform magnetic field was found by Case.[4] In the work that followed, the motion of particle spin in a strong magnetic field was studied in the Dirac representation.[8,9] It was in this representation that the dependence of the value of the anomalous magnetic moment on field strength was discovered (the dynamical nature of the anomalous magnetic moment).[10] However, a complete description of polarization in the FW representation has an advantage: it yields a simpler polarization operator.

Here we will derive an exact equation of motion for the spin of particles with an anomalous magnetic moment in a uniform magnetic field. The theory is based on the Dirac–Pauli equation in the Foldy–Wouthuysen representation. By comparing the motion of the particles and their spin we achieve an exact description of particle motion and the rotation, due to the anomalous magnetic moment, of the spin with respect to momentum.

Throughout the work we use the relativistic system of units $\hbar=c=1$.

2. DESCRIBING POLARIZATION IN THE FOLDY–WOUTHUYSEN REPRESENTATION

The description of polarization effects in the FW representation is simpler because the polarization operator in this representation reduces to the matrix $\mathbf{\Pi}=\beta\mathbf{\Sigma}$ (Refs. 11 and 12), where

$$\mathbf{\Pi}=\begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & -\boldsymbol{\sigma}\end{pmatrix},\quad \beta=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\quad \mathbf{\Sigma}=\begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & \boldsymbol{\sigma}\end{pmatrix},$$

with $\boldsymbol{\sigma}$ the Pauli matrix, and 0 and ± 1 standing for the respective 2-by-2 matrices. If to describe the state of a particle we use only one spinor, $\boldsymbol{\sigma}$ is the polarization operator. In the Dirac representation, the polarization operator of the particle is given by a more cumbersome expression.[9,11–16]

The operator equation of spin motion is found by calculating the commutator of the Hamiltonian operator and the polarization operator:

$$\frac{d\mathbf{\Pi}}{dt}=i[\mathscr{H},\mathbf{\Pi}]. \tag{1}$$

Since the absolute value of the polarization vector does not change with the passage of time, Eq. (1) can be written as

$$\frac{d\mathbf{\Pi}}{dt}=\frac{1}{2}(\mathbf{\Omega}\times\mathbf{\Pi}-\mathbf{\Pi}\times\mathbf{\Omega}),\quad \mathbf{\Omega}=\begin{pmatrix}\boldsymbol{\omega}' & 0\\ 0 & \boldsymbol{\omega}''\end{pmatrix}, \tag{2}$$

where $\mathbf{\Omega}$, $\boldsymbol{\omega}'$, and $\boldsymbol{\omega}''$ are the operators of the angular velocity of spin precession for the bispinor $\Psi=\binom{\psi}{\zeta}$ and the spinors ψ and ζ, respectively. The operator $\mathbf{\Omega}$ is quasidiagonal in view of the separation of the equations for the two spinors. Note that the operator $d\mathbf{\Pi}/dt$ is Hermitian.

There exists a certain arbitrariness in determining the operator $\mathbf{\Omega}$ from Eq. (2). In particular, the equation does not change under transformations of the form

$$\mathbf{\Omega}\rightarrow\mathbf{\Omega}'=\mathbf{\Omega}+\Lambda\mathbf{G}(\mathbf{\Sigma}\cdot\mathbf{G}), \tag{3}$$

where Λ is a scalar operator, $[\Lambda,\mathbf{\Sigma}]=[\Lambda,\mathbf{\Omega}]=0$, and $\mathbf{G}$ is a constant vector.

The operator $\mathbf{\Omega}$ can be used for a rigorous quantum mechanical description of spin motion,[17] since the average values of the projections of this operator

$$\langle\Omega_i\rangle=\int\Psi^\dagger\Omega_i\Psi dV,\quad i=x,y,z, \tag{4}$$

determine the spin rotation angles Φ_i about the ith axis: $\langle\Omega_i\rangle=d\Phi_i/dt$. Of course, it is much more practical to use only the upper spinor, and (4) assumes the form

$$\langle\omega_i'\rangle=\frac{d\Phi_i}{dt}=\int\psi^\dagger\omega_i'\psi dV, \tag{5}$$

with the functions ψ normalized to unity.

3. THE HAMILTONIAN OPERATOR IN THE FOLDY–WOUTHUYSEN REPRESENTATION

For a particle with an anomalous magnetic moment placed in a constant uniform magnetic field, the energy spectrum and the eigen-wave-functions were first found in the Dirac representation.[18] The same problem was successfully solved in the FW representation.[6,7] By applying a unitary transformation we can transform the Hamiltonian operator to the form[7]

$$\mathscr{H}=\beta\sqrt{\boldsymbol{\pi}_\perp^2+m^2-e\mathbf{\Sigma}\cdot\mathbf{H}}-\mu'\mathbf{\Pi}\cdot\mathbf{H}+\alpha_z p_z, \tag{6}$$

$$\alpha_z=\begin{pmatrix}0 & \sigma_z\\ \sigma_z & 0\end{pmatrix},$$

where $\mathbf{p}\equiv-i\nabla$ and $\boldsymbol{\pi}=\mathbf{p}=e\mathbf{A}$ are the operators of momentum and kinetic momentum, and $\mathbf{A}$ is the vector potential of the external field, which for a uniform field can be written as $\mathbf{A}=(\mathbf{H}\times\mathbf{r})/2$.

We assume that the magnetic field is directed along the z axis. The Hamiltonian $\mathscr{H}$ is not strictly diagonal, since the matrix α_z is not diagonal. However, $\mathscr{H}$ is diagonal when the particle is moving transversely, i.e., when the eigenvalue of the operator p_z is zero. Since $\partial\mathbf{A}/\partial z=0$, p_z commutes with $\mathscr{H}$, and this particular case is meaningful. The case of transverse particle motion is most interesting from the practical viewpoint, and when the particle moves in an arbitrary manner, we can always select a reference frame in which $\pi_z\Psi=p_z\Psi=0$. Then $\boldsymbol{\pi}_\perp=\boldsymbol{\pi}$, and the Hamiltonian operator is diagonal and has the form

$$\mathscr{H}=\beta\sqrt{\boldsymbol{\pi}^2+m^2-e\mathbf{\Sigma}\cdot\mathbf{H}}-\mu'\mathbf{\Pi}\cdot\mathbf{H}. \tag{7}$$

The fact that the Hamiltonian is diagonal indicates that the Hamiltonian corresponds to the FW representation. The Hamiltonian can be obtained in another way: by multiplying the Dirac–Pauli equation for a particle in a magnetic field,

$$[\gamma^\mu\pi_\mu-m+\mu'\mathbf{\Sigma}\cdot\mathbf{H}]\Psi=0,\quad \pi_\mu=p_\mu-eA_\mu,$$

into the operator $\gamma^\mu\pi_\mu+m+\mu'\mathbf{\Sigma}\cdot\mathbf{H}$. Using this approach, similar to the one used in Ref. 19, we can reduce the given equation to a form quadratic in the operator $\mathscr{E}\equiv i(\partial/\partial t)$ and then do the operator extraction of the square root.

If we use the weak-field approximation ($|e|H\ll m^2$), the Hamiltonian (7) corresponds to the one found by Suttorp and deGroot[20] (see also Ref. 21), and if $\mu'=0$, we arrive at the Hamiltonian of a Dirac particle obtained by Case.[4] In contrast to the Dirac wave function, the eigen-wave-function Ψ of the Hamiltonian (7) is the eigenfunction of the operator Π_z (see Ref. 19). These arguments suggest that Eq. (7) defines the Hamiltonian operator in the FW representation.

4. EXACT QUANTUM MECHANICAL EQUATION OF SPIN MOTION

The first to obtain the exact equation of spin motion for Dirac particles in a constant uniform magnetic field,

$$\frac{d\mathbf{\Sigma}}{dt}=\beta(E_+-E_-)\frac{\mathbf{\Sigma}\times\mathbf{H}}{H},\quad E_\pm=\sqrt{\boldsymbol{\pi}^2+m^2\pm eH}, \tag{8}$$

was Case (Ref. 4).[1)] The use of the operator $\mathbf{\Sigma}$ instead of $\mathbf{\Pi}$ to describe particle polarization is possible because the matrices β and $\mathbf{\Sigma}$ commute. The angular velocity of spin precession for Dirac particles is

$$\mathbf{\Omega}_D=-\beta(E_+-E_-)\frac{\mathbf{H}}{H}. \tag{9}$$

The equation of spin motion for particles with an anomalous magnetic moment can be found in a similar way. Linearizing the expression (7) for the Hamiltonian by the method used in Ref. 22 and calculating the commutator of the Hamiltonian with the polarization operator, we find that the exact operator equation of spin motion has the form (2), where the operator of the angular velocity of precession is defined as

$$\mathbf{\Omega}=-\beta\left[\frac{E_+-E_-}{H}+2\mu'\right]\mathbf{H}=\mathbf{\Omega}_D-2\beta\mu'\mathbf{H}. \tag{10}$$

The operator $\mathbf{\Omega}$ commutes with the Hamiltonian (7) and its eigenvalues for stationary states are

$$\mathbf{\Omega}=-\beta\left[\left(\sqrt{m^2+(2n+1)|e|H}+eH\right.\right.$$
$$\left.\left.-\sqrt{m^2+(2n+1)|e|H-eH}\right)\frac{1}{H}+2\mu'\right]\mathbf{H}. \tag{11}$$

Within the limits of the Dirac–Pauli equation, formula (11) is the exact expression for the frequency of precession (and hence of the rotation angles) of the spin of a spin-$\frac{1}{2}$ particle in a strong magnetic field. The formulas (10) and

(11) constitute a generalization of the Bargmann–Michel–Telegdi (BMT) equation[23] to the case where strong-field effects are taken into account.

As noted in Sec. 2, a term of the form $\Lambda\mathbf{G}(\boldsymbol{\Sigma}\cdot\mathbf{G})$, which does not change the equation of spin motion, must be added to the operator $\boldsymbol{\Omega}$. If $\mathbf{G}=\mathbf{H}$ holds, the presence of such a term may create the illusion that the angular velocity of spin precession depends on the sign of the projection of spin on the direction of the magnetic field, since the operator $\boldsymbol{\Sigma}\cdot\mathbf{H}=\Sigma_z H$ commutes with the Hamiltonian. In reality, however, the main characteristic of spin motion is the equation of spin motion, i.e., the expression for the operator $d\boldsymbol{\Pi}/dt$ rather than for the operator $\boldsymbol{\Omega}$, which cannot be defined unambiguously. This suggests that the average value of the operator $\boldsymbol{\Omega}$ calculated by (2) and (4) determines the average velocity of spin rotation if and only if this operator commutes with $\boldsymbol{\Sigma}$ and $\boldsymbol{\Pi}$ and hence contains no spin matrices. Otherwise, when we pass to the semiclassical approximation by averaging the operators in Eq. (2) for the spin motion, the operator $\boldsymbol{\Omega}$, in accordance with this equation, cannot be regarded as acting directly on the wave function Ψ, so that for it formula (4) is invalid. Hence, only the operator $\boldsymbol{\Omega}$ that is defined by (10) and (11) and commutes with $\boldsymbol{\Sigma}$ and $\boldsymbol{\Pi}$ makes it possible to find an expression for the average angular rate of spin rotation.

5. EXACT QUANTUM MECHANICAL DESCRIPTION OF PARTICLE MOTION

Studying the motion of a particle is more complicated than studying the motion of spin, since the operator of the angular velocity of a particle moving along a circular orbit in a magnetic field does both commute with the kinetic-momentum operator. Baĭer *et al.*[16] found the equation of motion for a scalar particle in a magnetic field solvable by the method of successive approximations, and Suttorp and deGroot[20] found the equation of motion for a spin-$\frac{1}{2}$ particle in the weak-field approximation. Here we will show that it is possible to achieve an exact quantum mechanical description of the motion of a spin-$\frac{1}{2}$ particle by comparing particle motion and spin motion.

It is natural to find the angular velocity operator $\boldsymbol{\omega}$ by solving an equation similar to (2):

$$\frac{d\boldsymbol{\pi}}{dt}=\frac{1}{2}(\boldsymbol{\omega}\times\boldsymbol{\pi}-\boldsymbol{\pi}\times\boldsymbol{\omega}). \tag{12}$$

However, $\boldsymbol{\omega}$ is not an operator whose average value is the angular velocity of particle rotation. Since the operators $\boldsymbol{\pi}$ and $\boldsymbol{\omega}$ do not commute, when we pass to the semiclassical approximation by averaging the operators, $\boldsymbol{\omega}$ cannot be consider an operator acting directly on the wave function Ψ. Hence $\langle\omega_i\rangle=d\phi_i/dt$, where ϕ_i are the angles of rotation of the momentum vector about the axes $i=x,y,z$. Consequently, although equations of the form (12) provide a meaningful description of particle motion within the accuracy to which they were derived, by themselves they do not yield expressions for the angular velocity of particle motion. At the same time, when the particle moves in a constant uniform magnetic field, its angular velocity has a fixed value, which is proved by the following arguments. It is convenient to write the Hamiltonian operator (7) in the form

$$\begin{aligned}&\mathscr{H}=\mathscr{H}_1+\mathscr{H}_2,\\&\mathscr{H}_1=\beta\sqrt{\boldsymbol{\pi}^2+m^2-e\boldsymbol{\Sigma}\cdot\mathbf{H}}=\beta\sqrt{m^2+(\boldsymbol{\Sigma}\cdot\boldsymbol{\pi})^2},\\&\mathscr{H}_2=-\mu'\boldsymbol{\Pi}\cdot\mathbf{H}.\end{aligned} \tag{13}$$

Here $\mathscr{H}_1$ is the Dirac Hamiltonian. It commutes with the operator $\boldsymbol{\Pi}\cdot\boldsymbol{\pi}=\beta\boldsymbol{\Sigma}\cdot\boldsymbol{\pi}$, since $\beta\boldsymbol{\Sigma}=\boldsymbol{\Sigma}\beta$, and with the operator $|\boldsymbol{\pi}|=\sqrt{\boldsymbol{\pi}^2}$. Allowing for the fact that $\boldsymbol{\Pi}\cdot\boldsymbol{\pi}=\Pi_\pi|\boldsymbol{\pi}|$, we find that

$$[\mathscr{H}_1,\Pi_\pi|\boldsymbol{\pi}|]=[\mathscr{H}_1,\Pi_\pi]|\boldsymbol{\pi}|=0,\quad [\mathscr{H}_1,\Pi_\pi]=0,$$

and for Dirac particles the projection of the polarization operator on the direction of the kinetic momentum Π_π is conserved. This result was obtained by Sokolov and Ternov[13] and Baĭer *et al.*,[16] who used the Dirac representation. The angular velocities of the gyration of a Dirac particle in a circular orbit and of the precession of the particle's spin coincide because the projection of the polarization operator on the direction of the kinetic momentum is conserved.

We introduce the operator of the angular velocity of rotational motion, $\mathbf{o}$, whose average, in contrast to the average of $\boldsymbol{\omega}$, reflects the real particle motion, $\langle o_i\rangle=d\phi_i/dt$. The above implies that $\mathbf{o}=\boldsymbol{\Omega}_D$. Since

$$[\mathscr{H}_2,\boldsymbol{\pi}]=0,\quad \frac{d\boldsymbol{\pi}}{dt}=i[\mathscr{H},\boldsymbol{\pi}]=i[\mathscr{H}_1,\boldsymbol{\pi}],$$

the equation of motion of a particle with an anomalous magnetic moment does not differ from the equation of motion of a Dirac particle. Hence the operator $\mathbf{o}$ for such a particle is defined by the exact expression

$$\mathbf{o}=\boldsymbol{\Omega}_D=-\beta(E_+-E_-)\,\frac{\mathbf{H}}{H}. \tag{14}$$

The rotation of spin with respect to the momentum vector is given by the difference of the angular velocities of particle spin precession, $\boldsymbol{\Omega}$, and of the particle's orbital motion, $\mathbf{o}$. The exact expression for this quantity can be found from Eqs. (10) and (14):

$$\boldsymbol{\Omega}-\mathbf{o}=-2\beta\mu'\mathbf{H}. \tag{15}$$

6. EQUATION OF PARTICLE MOTION IN A STRONG MAGNETIC FIELD

Although in Sec. 5 we derived an exact description of the motion of the particle with an anomalous magnetic moment in a uniform magnetic field by introducing the operator $\mathbf{o}$, it would be interesting to find the equations of motion of a particle in a strong magnetic field. The derivation of a theory of strong-field effects requires stepping outside the scope of the Dirac–Pauli equation. However, to calculate the corrections introduced by such effects into the equations of motion of particle and spin correctly we must first derive these equations from the Dirac–Pauli equation. Hence in this section we will find an equation of particle motion based on the Dirac–Pauli equation and valid to within terms of order H^4.

It is impossible to derive an exact equation, since the operator $\boldsymbol{\pi}$ commutes neither with $\boldsymbol{\pi}^2$ nor with the kinetic-energy operator $\epsilon=\sqrt{\boldsymbol{\pi}^2+m^2}$:

$$[\boldsymbol{\pi}^2,\boldsymbol{\pi}]=ie(\mathbf{H}\times\boldsymbol{\pi}-\boldsymbol{\pi}\times\mathbf{H})=2ie\mathbf{H}\times\boldsymbol{\pi}\neq 0,$$

$$[\epsilon,\boldsymbol{\pi}]\neq 0.$$

The equation of particle motion written for the kinetic-momentum operator $\boldsymbol{\pi}$ has the form (12). We wish to show that the expression for $\boldsymbol{\omega}$ does not change when we establish the equations of motion for two other operators: the identity operator in the direction of the motion,

$$\frac{1}{2}\left(\frac{1}{|\boldsymbol{\pi}|}\,\boldsymbol{\pi}+\boldsymbol{\pi}\,\frac{1}{|\boldsymbol{\pi}|}\right),$$

and the velocity operator, which can be written as

$$\mathbf{v}=\frac{1}{2}\left(\frac{1}{\epsilon}\,\boldsymbol{\pi}+\boldsymbol{\pi}\,\frac{1}{\epsilon}\right).$$

The reason is that both $|\boldsymbol{\pi}|$ and ε commute with the Hamiltonian $\mathscr{H}$ and, as we will see shortly, with the operator $\boldsymbol{\omega}$. For any operator A that commutes with $\mathscr{H}$ and $\boldsymbol{\omega}$ we have

$$\frac{d}{dt}(A\boldsymbol{\pi})=iA[\mathscr{H},\boldsymbol{\pi}]=\frac{1}{2}A(\boldsymbol{\omega}\times\boldsymbol{\pi}-\boldsymbol{\pi}\times\boldsymbol{\omega})$$

$$=\frac{1}{2}(\boldsymbol{\omega}\times A\boldsymbol{\pi}-A\boldsymbol{\pi}\times\boldsymbol{\omega}),$$

where $\boldsymbol{\omega}$ is defined by Eq. (12). A similar relationship can easily be obtained for $d(\boldsymbol{\pi}A)/dt$, which makes it possible to prove the above assertion.

We can easily derive the exact relationships that make it possible to express the commutator of two arbitrary operators A and B in terms of the commutator of A^2 and B:[2)]

$$[A^{-1},B]=A^{-1}[B,A]A^{-1}, \tag{16}$$

$$[A,B]=\frac{1}{4}\{A^{-1},[A^2,B]\}_+-\frac{1}{4}[[A,[A,B]],A^{-1}], \tag{17}$$

where $A^{-1}\equiv 1/A$, and $\{\cdots,\cdots\}_+$ is an anticommutator.

Equation (17) makes it possible to find $[A,B]$ by the method of successive approximations if the commutator of the operator is small in comparison to the product of the operators. In the first approximation,

$$[A,B]\approx\frac{1}{4}\{A^{-1},[A^2,B]\}_+$$

$$-\frac{1}{16}[[A,\{A^{-1},[A^2,B]\}_+],A^{-1}].$$

Since $[A,\{A^{-1},C\}_+]=A^{-1}[A^2,C]A^{-1}$ for any operator C, we have

$$[A,B]\approx\frac{1}{4}\{A^{-1},C\}_+-\frac{1}{16}[A^{-1}[A^2,C]A^{-1},A^{-1}]$$

$$\approx\frac{1}{4}\{A^{-1},C\}_+-\frac{1}{64}A^{-2}\{A^{-1},[A^2,[A^2,C]]\}_+A^{-2}, \tag{18}$$

where $C=[A^2,B]$.

If we use Eqs. (16)–(18), we can find the operator $\boldsymbol{\omega}$ to within terms of order H^4. Next we introduce the operator $\mathscr{A}=\sqrt{\boldsymbol{\pi}^2+m^2-e\boldsymbol{\Sigma}\cdot\mathbf{H}}$. Then for the Hamiltonian (7) we have $[\mathscr{H},\boldsymbol{\pi}]=\beta[\mathscr{A},\boldsymbol{\pi}]$. The commutators in Eq. (18) are

$$[\mathscr{A}^2,\boldsymbol{\pi}]=[\boldsymbol{\pi}^2,\boldsymbol{\pi}]=2ie\mathbf{H}\times\boldsymbol{\pi},$$

$$[\mathscr{A}^2,[\mathscr{A}^2,[\mathscr{A}^2,\boldsymbol{\pi}]]]=8ie^3H^2\mathbf{H}\times\boldsymbol{\pi}. \tag{19}$$

Combining (12) with (16)–(19), we arrive at an equation describing particle motion:

$$\frac{d\boldsymbol{\pi}}{dt}=-\beta\,\frac{e}{2}\{\mathscr{A}^{-1},\mathbf{H}\times\boldsymbol{\pi}\}_+$$

$$+\frac{1}{8}\beta e^3H^2\mathscr{A}^{-2}\{\mathscr{A}^{-1},\mathbf{H}\times\boldsymbol{\pi}\}_+\mathscr{A}^{-2}.$$

To within a double commutator of order H^5,

$$\frac{d\boldsymbol{\pi}}{dt}=-\beta\,\frac{e}{2}\left\{\frac{1}{\mathscr{A}}\left(1-\frac{e^2H^2}{4\mathscr{A}^4}\right),\mathbf{H}\times\boldsymbol{\pi}\right\}_+. \tag{20}$$

According to Eqs. (12) and (20), the expression for the operator $\boldsymbol{\omega}$ is

$$\boldsymbol{\omega}=-\beta\,\frac{e}{\mathscr{A}}\left(1-\frac{e^2H^2}{4\mathscr{A}^4}\right)\mathbf{H}. \tag{21}$$

The corrections to Eqs. (20) and (21) are of order H^5. Hence Eqs. (20) and (21) are the equation of particle motion and the expression for the angular velocity of the particle's orbital motion to within terms of order H^4 inclusive. Note that both (20) and (21) demonstrate the validity of the above assumption that the operator $\boldsymbol{\omega}$ commutes with $|\boldsymbol{\pi}|=\sqrt{\boldsymbol{\pi}^2}$ and ϵ.

Clearly, for Dirac particles the equations of particle and spin motion have same structure. The commutators corresponding to (19) are

$$[\mathscr{A}^2,\boldsymbol{\Pi}]=-e[(\boldsymbol{\Sigma}\cdot\mathbf{H}),\boldsymbol{\Pi}]=2ie\mathbf{H}\times\boldsymbol{\Pi},$$

$$[\mathscr{A}^2,[\mathscr{A}^2,[\mathscr{A}^2,\boldsymbol{\Pi}]]]=8ie^3H^2\mathbf{H}\times\boldsymbol{\Pi}.$$

As a result the approximate (to within terms of order H^4 inclusive) equation of spin motion for Dirac particles can be written as

$$\frac{d\boldsymbol{\Pi}}{dt}=-\beta\,\frac{e}{2}\left\{\frac{1}{\mathscr{A}}\left(1-\frac{e^2H^2}{4\mathscr{A}^4}\right),\mathbf{H}\times\boldsymbol{\Pi}\right\}_+. \tag{22}$$

The fact that Eqs. (20) and (22) are identical in form supports the conclusion that the angular velocities of the orbital motion of a Dirac particle and of the precession of particle spin coincide. Equation (28) can be transformed to a form that agrees with the exact equation (8).

7. DISCUSSION AND CONCLUSIONS

We have carried out an exact quantum mechanical description of the motion of spin-$\frac{1}{2}$ particles with an anomalous magnetic moment and the motion of spin in a uniform magnetic field. Since we used the Dirac–Pauli equation in the Foldy–Wouthuysen representation, we would like to know

how these equations are modified when we go over to other representations. From the general ideas of the theory of representations it follows that in this case the operators undergo a unitary transformation but the form of the equations does not change. However, it is clear that if there is an external field, such a transformation changes not only the form of the polarization operators **O**, and coordinates, x, y, and z, but also of the operators of momentum, $\mathbf{p}\equiv -i\nabla$, and kinetic momentum, $\boldsymbol{\pi}=\mathbf{p}-e\mathbf{A}$. In particular, for this reason the dependence of the equation of spin motion in the Dirac representation on the operator $\boldsymbol{\pi}'=U^{-1}\boldsymbol{\pi}U$, with U the operator of the unitary transformation, is exactly the same as the dependence of that equation in the FW representation on the operator $\boldsymbol{\pi}$. However, these two equations differ in their dependence on $\boldsymbol{\pi}$. As a result one is tricked into thinking that the nature of the spin motion has changed or that new effects have emerged.

As is known, the FW representation is special because in it the operators are even and can be assigned to the corresponding classical quantities. The other representations do not possess this property. The equation of spin motion in a uniform magnetic field derived in the present paper can be transformed into the Dirac representation by using a unitary-transformation operator, which for the given case was found by Tsai.[7]

An exact quantum mechanical description of the motion of particle spin with respect to particle momentum in a strong magnetic field is important from the practical viewpoint. Since the particle moves along a circular orbit, the angular velocities of rotation of the particle and the particle spin can be compared, with the difference of these two quantities yielding the rotation of the spin in relation to the particle momentum. Dirac particles do not experience such motion ($\boldsymbol{\Omega}_D=\mathbf{o}$), and the rotation of spin with respect to momentum occurs only for particles with an anomalous magnetic moment. Using Eqs. (4) and (15), we can determine this angle exactly. For instance, if initially the particle beam was polarized along the momentum vector, after time t has elapsed the polarization vector of the beam forms is directed at an angle θ to the momentum vector, with

$$\theta=(\langle\Omega_z\rangle-\langle o_z\rangle)t=-2\mu' Ht. \tag{23}$$

Of course, we must bear in mind that Eqs. (15) and (23) are exact only within the framework of the Dirac–Pauli equation. As is known, allowance for strong-field effects provides corrections both to the anomalous magnetic moment of the electron and to the Dirac–Pauli equation (see Ref. 9). There are also radiative corrections to the equation of spin motion, which allow for damping.[16] The detection of corrections to Eqs. (10), (15), and (23) in experiments would be a clear indication that the limits of the Dirac–Pauli have been passed,[3)] and establishing the values of these corrections would make it possible to demonstrate the extent to which the theoretical results agree with the experimental data.

According to the method of exact solutions formulated in Refs. 9, 13, and 15, the exact solutions of the Dirac and Dirac–Pauli equations are taken as the zeroth approximation, and the corrections to these solutions are calculated by perturbation-theory techniques by abandoning the single-particle approach. This makes it possible to use the results obtained in this paper if strong-field effects must be taken into account more thoroughly.

*)E-mail: silenko@inp.minsk.by

1)Note that formula (8) and the other formulas in this section are exact in the sense that they are equivalent to the initial Dirac equation. However, the Dirac and Dirac–Pauli equations obtained in the single-particle approximation do not give an exhaustive description of the state of a particle. In particular, they do not allow for radiative corrections.

2)A similar method was used in Ref. 16.

3)In particular, the linear dependence between θ and H follows rigorously from the Dirac–Pauli equation.

[1] W. Pauli, Rev. Mod. Phys. **13**, 203 (1941).
[2] V. G. Bagrov, D. M. Gitman, I. M. Ternov *et al.*, *Exact solutions of Relativistic Wave Equations*, Kluwer Academic Publishers, Dordrecht, Boston (1990) [Russian orig., Novosibirsk, Nauka (1982)].
[3] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).
[4] K. M. Case, Phys. Rev. **95**, 1323 (1954).
[5] E. Eriksen, Phys. Rev. **111**, 1011 (1958).
[6] Tsai Wu-Yang and A. Yildiz, Phys. Rev. D **4**, 3643 (1971).
[7] Tsai Wu-Yang, Phys. Rev. D **7**, 1945 (1973).
[8] I. M. Ternov, Zh. Éksp. Teor. Fiz. **98**, 1169 (1990) [Sov. Phys. JETP **71**, 654 (1990)].
[9] I. M. Ternov and O. F. Dorofeev, Fiz. Élem. Chastits At. Yadra **25**, 5 (1994) [Phys. Part. Nuclei **25**, 1 (1994)].
[10] I. M. Ternov, V. Bagrov, V. A. Bordovitsyn, and O. F. Dorofeev, Zh. Éksp. Teor. Fiz. **55**, 2273 (1968) [Sov. Phys. JETP **28**, 1206 (1969)].
[11] D. M. Fradkin and R. H. Good, Rev. Mod. Phys. **33**, 343 (1961).
[12] I. M. Ternov, V. R. Khalilov, and V. N. Rodionov, *Interaction of Charged Particles with an External Electromagnetic Field* [in Russian], Moscow Univ. Press, Moscow (1982).
[13] A. A. Sokolov and I. M. Ternov, *The Relativistic Electron* [in Russian], Nauka, Moscow (1983).
[14] I. M. Ternov and V. V. Mikhaĭlin, *Synchrotron Radiation* [in Russian], Énergoatomizdat, Moscow (1986).
[15] I. M. Ternov, V. Ch. Zhukovskiĭ, and A. V. Borisov, *Quantum Processes in a Strong External Field* [in Russian], Moscow Univ. Press, Moscow (1989).
[16] B. N. Baĭer, V. M. Katkov, and V. S. Fadin, *Emission of Radiation by Relativistic Electrons* [in Russian], Atomizdat, Moscow (1973).
[17] A. Ya. Silenko, Zh. Éksp. Teor. Fiz. **107**, 1240 (1995) [JETP **80**, 690 (1995)].
[18] I. M. Ternov, V. G. Bagrov, and V. Ch. Zhukovskiĭ, Vestnik Moskov. Univ. Ser. Fiz. Astronom. No. 1, 30 (1966).
[19] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed., Pergamon Press, Oxford (1991).
[20] L. G. Suttorp and S. R. deGroot, Nuovo Cimento A **65**, 245 (1970); S. R. deGroot and L. G. Suttorp, *Foundations of Electrodynamics*, North-Holland, Amsterdam (1972).
[21] A. Ya. Silenko, Teor. Mat. Fiz. **105**, 46 (1995).
[22] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[23] V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Lett. **2**, 435 (1959).

Translated by Eugene Yankovsky

The role of the polarization mechanism for emission of radiation by atoms over a broad photon frequency range

A. V. Korol'

Russian Marine Technical University, 198262 St. Petersburg, Russia

A. G. Lyalin*)

Research Physics Institute at the St. Petersburg State University, 198904 St. Petersburg, Russia

O. I. Obolenskiĭ and A. V. Solov'ev

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 198904 St. Petersburg, Russia
(Submitted 12 July 1997)
Zh. Éksp. Teor. Fiz. **114**, 458–473 (August 1998)

This paper develops an effective method for calculating the bremsstrahlung cross section with allowance for the polarization mechanism. We calculate the cross section of bremsstrahlung produced in the scattering of electrons and positrons by H and Kr atoms. We also demonstrate the important role of polarization bremsstrahlung in the formation of the total emission spectrum over the entire frequency range. © *1998 American Institute of Physics.* [S1063-7761(98)00708-2]

1. INTRODUCTION

When charged particles collide with a target that has an internal electron structure, bremsstrahlung arises as a result of two different mechanisms. The first is known as ''ordinary'' bremsstrahlung and occurs as a result of the braking of an incident particle by the static field of an atom (see, e.g., Refs. 1 and 2), while the second is known as ''polarization'' bremsstrahlung and occurs as a result of dynamic polarization of the target by the electric field of the incident particle (see, e.g., Refs. 3–5).

In this paper we study the role of polarization bremsstrahlung formed in the process of electron and positron scattering by atoms within a broad spectral range. An analysis of the relative role of the two bremsstrahlung mechanisms is carried out using the hydrogen and krypton atoms as examples.

The hydrogen atom is a one-electron system, and for such a system one can derive closed analytical expressions. The polarization bremsstrahlung mechanism that occurs as a result of collisions of electrons, positrons, and protons with the hydrogen atom in the ground state was studied by Buĭmistrov *et al.*[6,7] and Dubois and Maquet.[8] They found that this mechanism is important to the formation of the total bremsstrahlung spectrum. In the present paper we carry out a detailed analysis of the role that the polarization mechanism plays in collisions of charged particles with the hydrogen atom in an excited state. We find that there are substantial differences in the behavior of the bremsstrahlung cross section when the target is in its ground state and when it is in an excited state. Our calculations show that as the principal quantum number of the state of the electron in the hydrogen atom increases, the relative role of the polarization mechanism becomes less important because the screening of the nuclear charge by the electron weakens. However, the spectral dependence of the cross section of bremsstrahlung produced by scattering of charged particles by the hydrogen atom in an excited state exhibits singularities absent when the atom is in the ground state. The numerical calculations of the cross sections are done for the states $1s$, $2s$, and $3s$.

Prior to publication of our recent paper,[9] calculations of the cross section of polarization bremsstrahlung resulting from the collision of charged particles with multielectron atoms were carried only in a relatively narrow photon frequency range characteristic of excitation of giant resonances in atoms. The interest in this spectral range is due to the collective multielectron nature of giant resonances and the related large value of atomic polarizability, which imply that the polarization mechanism plays the leading role in the total bremsstrahlung spectrum.

The bremsstrahlung spectrum formed in electron–atom collisions has been calculated in the nonrelativistic Born approximation (BA) for the Ar atom by Amus'ya *et al.*[10–13] They found the total bremsstrahlung cross section in the photon frequency range characteristic of excitations of electrons belonging to the $3p$-subshell of Ar ($\omega = 10$–50 eV). Similar calculations have been done in the spectral range near the ionization potential of the $4d$-subshell ($\omega = 100$–200 eV) of the Xe and La atoms[14–16] in the Born approximation and for Xe, Ba, La, and Eu in the distorted partial wave approximation (DPWA).[17–22] It was found that in bremsstrahlung spectra the polarization mechanism gives rise to a resonance structure near the $3p$ ionization potential of Ar and the $4d$ ionizati on potential of Xe, Be, La, and Eu. The existence of such peaks was proved in experiments in which the emission spectra for the electron scattering by La and rare-earth elements,[23] Xe (Refs. 15 and 24), and Ba (Ref. 25) were measured.

In this paper we also study the bremsstrahlung process

for electrons with energies $\varepsilon_1 = 1$–25 keV scattered by the Kr atom. The spectral dependence of the total bremsstrahlung cross section (i.e., the cross section that incorporates the contributions of ordinary and polarization bremsstrahlung and an interference term) is calculated for the entire photon frequency range from $\omega = 10$ eV to ε_1. We find that the polarization mechanism plays an important role in forming the total bremsstrahlung spectrum not only near the ionization potentials of atomic multielectron subshells but also in the range of the emitted-photon frequencies. Allowance for the polarization bremsstrahlung mechanism alters dramatically the common idea that the cross section is a monotonic function of the emitted-photon frequency. The dependence of the total bremsstrahlung cross section of the photon energy is a nonmonotonic function with characteristic resonance singularities. When the polarization mechanism is taken into account, the nonmonotonic nature of the total bremsstrahlung cross section appears even in a one-electron system (the hydrogen atom or hydrogenlike ion), where no multiparticle effects are present.

Throughout the paper we use the atomic system of units ($|e| = \hbar = m_e = 1$).

2. TOTAL BREMSSTRAHLUNG AMPLITUDE AND CROSS SECTION

The total bremsstrahlung amplitude $f_{\rm tot}$ is the sum of the ordinary ($f_{\rm ord}$) and polarization ($f_{\rm pol}$) bremsstrahlung amplitudes:

$$f_{\rm tot} = f_{\rm ord} + f_{\rm pol}. \tag{1}$$

In the lowest order of the nonrelativistic perturbation theory in the electron–photon interaction and in the Coulomb interaction $V = 1/|\mathbf{r} - \mathbf{r}_a|$ between the incident ($\mathbf{r}$) and atomic ($\mathbf{r}_a$) electrons that excites the atom virtually, the amplitudes $f_{\rm ord}$ and $f_{\rm pol}$ are given by the expressions

$$f_{\rm ord} = \langle \mathbf{p}_2^{(-)} | \mathbf{e}\cdot\mathbf{r} | \mathbf{p}_1^{(+)} \rangle, \tag{2}$$

$$f_{\rm pol} = -\sum_n \left\{ \frac{\langle 0|\mathbf{e}\cdot\mathbf{D}|n\rangle\langle \mathbf{p}_2^{(-)}, n|V|\mathbf{p}_1^{(+)}, 0\rangle}{\omega_{n0} - \omega - i0} + \frac{\langle \mathbf{p}_2^{(-)}, 0|V|\mathbf{p}_1^{(+)}, n\rangle\langle n|\mathbf{e}\cdot\mathbf{D}|0\rangle}{\omega_{n0} + \omega} \right\}, \tag{3}$$

where $|\mathbf{p}_1^{(+)}\rangle$ and $|\mathbf{p}_2^{(-)}\rangle$ are the wave functions of the incident and scattered electrons with momenta $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively. The ''plus'' and ''minus'' superscripts denote the solutions of the Schrödinger equation that asymptotically are, respectively, a diverging (plus) and converging (minus) wave at infinity. The partial-wave expansion of the wave function of the incident particle in the initial ($j = 1$) and final ($j = 2$) states is

$$|\mathbf{p}_j^{(\pm)}\rangle = 4\pi \sqrt{\frac{\pi}{p_j}} \sum_{l,m} i^l \exp[\pm i\delta_l(p_j)] \frac{P_{\nu_j}(r)}{r} \times Y_{lm}^*(\hat{\mathbf{p}}_j) Y_{lm}(\hat{\mathbf{r}}). \tag{4}$$

Here $\hat{\mathbf{a}} \equiv \mathbf{a}/|\mathbf{a}|$, where $\mathbf{a}$ is any one of the $\mathbf{p}_j$ or $\mathbf{r}$, $\delta_l(p)$ is the phase shift, and the symbol ν stands for the set of quantum numbers (p,l). The radial wave function $P_\nu(r)$ satisfies the Schrödinger equation with a ''frozen'' atomic core and is normalized to energy in rydbergs.

The operator $\mathbf{e}\cdot\mathbf{D}$ in (3) is the operator of the dipole interaction of an atomic electron and the electromagnetic field, ω and $\mathbf{e}$ are the electron energy and polarization vector, $\omega_{n0} = E_n - E_0$ is the energy of an atomic transition from the initial state $|0\rangle$ to an excited state $|n\rangle$ in the discrete or continuous spectrum.

The amplitude (3) does not take into account the exchange between the incident and atomic electrons. In the present paper we assume that the incident electron energy before and after the emission process is large compared to the ionization potentials of the atomic shells that for a given photon frequency ω provide the main contribution to the sum over the excited states.

Following Ref. 18, we express the amplitude (3) for a spherically symmetric atom in terms of the generalized dynamic dipole polarizability $\alpha(\omega, Q)$. To this end we write the Coulomb interaction operator V in the form of a Fourier integral:

$$V \equiv \frac{1}{|\mathbf{r} - \mathbf{r}_a|} = \frac{1}{2\pi^2} \int \frac{d\mathbf{Q}}{Q^2} \exp\{-i\mathbf{Q}\cdot(\mathbf{r} - \mathbf{r}_a)\}. \tag{5}$$

Inserting (5) into (3) yields

$$f_{\rm pol} = -\frac{i}{2\pi^2} \int d\mathbf{Q} \frac{\mathbf{e}\cdot\mathbf{Q}}{Q^2} \langle \mathbf{p}_2^{(-)} | e^{-i\mathbf{Q}\cdot\mathbf{r}} | \mathbf{p}_1^{(+)} \rangle \alpha(\omega, Q), \tag{6}$$

where

$$i(\mathbf{e}\cdot\mathbf{q})\alpha(\omega, Q) = \sum_n \left\{ \frac{\langle 0|\mathbf{e}\cdot\mathbf{D}|n\rangle\langle n|e^{i\mathbf{Q}\cdot\mathbf{r}_a}|0\rangle}{\omega_{n0} - \omega - i0} + \frac{\langle 0|e^{i\mathbf{Q}\cdot\mathbf{r}_a}|n\rangle\langle n|\mathbf{e}\cdot\mathbf{D}|0\rangle}{\omega_{n0} + \omega} \right\}. \tag{7}$$

Integration in (6) is done over the entire space of the vectors $\mathbf{Q}$. At $Q = 0$ the generalized dynamic polarizability $\alpha(\omega, Q)$ becomes the ordinary dynamic polarizability $\alpha(\omega)$.

The representation of $f_{\rm pol}$ in the form (6) makes it easy to pass to the Born limit in the amplitude of polarization bremsstrahlung. Indeed, replacing the distorted waves $|\mathbf{p}_{1,2}^{(\pm)}\rangle$ in (6) with the free-particle wave functions $|\tilde{\mathbf{p}}_{1,2}^{(\pm)}\rangle = \exp(i\mathbf{p}_{1,2}\cdot\mathbf{r})$ and bearing in mind that

$$\langle \tilde{\mathbf{p}}_2^{(-)} | \exp(-i\mathbf{Q}\cdot\mathbf{r}) | \tilde{\mathbf{p}}_1^{(+)} \rangle = (2\pi)^3 \delta(\mathbf{q} - \mathbf{Q}),$$

we arrive at an expression for $f_{\rm pol}$ in the Born approximation:

$$f_{\rm pol}^B = -4\pi i \frac{\mathbf{e}\cdot\mathbf{q}}{q^2} \alpha(\omega, q), \tag{8}$$

where $\mathbf{q} = \mathbf{p}_1 - \mathbf{p}_2$ is the momentum transferred to the atom in the collision process.

The expression (8) for the amplitude of polarization bremsstrahlung in the Born approximation and the corresponding expression for the bremsstrahlung cross section were first obtained by Amus'ya *et al.*[10]

Substituting (4) in (2) and (6), we arrive at expressions for $f_{\rm ord}$ and $f_{\rm pol}$ in the form of sums of partial contributions. The structure of both series is the same and is given by the expression

$$f=\frac{16\pi^3}{\sqrt{p_1p_2}}\sqrt{\frac{4\pi}{3}}\sum_{l_1,l_2} i^{l_1-l_2}\exp[i(\delta_{l_1}(p_1)+\delta_{l_2}(p_2))] \times(-1)^{l_>}\sqrt{l_>}\,T_{l_2l_1}(\hat{\mathbf{e}},\hat{\mathbf{p}}_1,\hat{\mathbf{p}}_2)R_{l_2l_1}, \tag{9}$$

where $l_2=l_1\pm1$ in accordance with dipole selection rules, and $l_>=\max\{l_1,l_2\}$. The factor $T_{l_2l_1}(\hat{\mathbf{e}},\hat{\mathbf{p}}_1,\hat{\mathbf{p}}_2)$ depends on the angular variables of the vectors $\mathbf{e}$, $\mathbf{p}_1$, and $\mathbf{p}_2$:

$$T_{l_2l_1}(\hat{\mathbf{e}},\hat{\mathbf{p}}_1,\hat{\mathbf{p}}_2)=\sum_{m_1,m_2,\mu}(-1)^{m_2}\begin{pmatrix} l_2 & 1 & l_1\\ -m_2 & \mu & m_1\end{pmatrix} \times Y^*_{1\mu}(\hat{\mathbf{e}})Y^*_{l_1m_1}(\hat{\mathbf{p}}_1)Y_{l_2m_2}(\hat{\mathbf{p}}_2). \tag{10}$$

In (9), the function

$$R_{l_2l_1}=\begin{cases}+R^{\rm ord}_{l_2l_1} & \text{for ordinary bremsstrahlung,}\\ -R^{\rm pol}_{l_2l_1} & \text{for polarization bremsstrahlung,}\end{cases}$$

can be expressed in terms of the integrals

$$R^{\rm ord}_{l_2l_1}=\langle\nu_2\|r\|\nu_1\rangle, \tag{11}$$

$$R^{\rm pol}_{l_2l_1}=\frac{2}{\pi}\int_0^\infty dQ\,Q\langle\nu_2\|j_1(Qr)\|\nu_1\rangle\alpha(\omega,Q), \tag{12}$$

where $j_1(Qr)$ is the spherical Bessel function.

The differential cross section of total bremsstrahlung, which characterizes the spectral distribution of the emitted radiation, can be written as

$$\omega\left(\frac{d\sigma}{d\omega}\right)=\frac{1}{(2\pi)^4}\frac{\omega^4}{c^3}\frac{p_2}{p_1}\int d\Omega_{\mathbf{p}_2}\,d\Omega_\gamma\sum_\lambda|f_{\rm tot}|^2, \tag{13}$$

where c is the speed of light. The integration in (13) is over the directions of propagation of the scattered electron ($d\Omega_{\mathbf{p}_2}$) and the emitted photon ($d\Omega_\gamma$), and summation is over the photon polarizations (λ).

Plugging (9)–(12) into (13), we arrive at an expression for the total bremsstrahlung cross section:

$$\omega\left(\frac{d\sigma}{d\omega}\right)=\frac{32\pi^2}{3}\frac{\omega^4}{c^3p_1^2}\sum_{l_1,l_2}l_>|R^{\rm ord}_{l_2l_1}-R^{\rm pol}_{l_2l_1}|^2. \tag{14}$$

Since bremsstrahlung is formed by two different mechanisms, it is natural to write the total cross section (13) as the sum of the cross sections of ordinary bremsstrahlung and polarization bremsstrahlung and an interference term:

$$\omega\left(\frac{d\sigma}{d\omega}\right)=\omega\left(\frac{d\sigma}{d\omega}\right)_{\rm ord}+\omega\left(\frac{d\sigma}{d\omega}\right)_{\rm pol}+\omega\left(\frac{d\sigma}{d\omega}\right)_{\rm int}. \tag{15}$$

The expression for the cross section of ordinary bremsstrahlung, $\omega(d\sigma/d\omega)_{\rm ord}$ can be derived from (14) if we put $R^{\rm pol}_{l_2l_1}=0$. If in (14) we put $R^{\rm ord}_{l_2l_1}=0$, we arrive at the expression for the cross section of polarization bremsstrahlung, $\omega(d\sigma/d\omega)_{\rm pol}$.

Calculating the generalized dynamic dipole polarizability of an excited hydrogen atom by the Coulomb Green's function method

Equations (9)–(13) show that the behavior of the total bremsstrahlung cross section is determined by the generalized dynamic polarizability of the target atom. The generalized polarizability describes the dynamic response of the target to the external field of the incident particle. For multielectron atoms such a response is usually determined by multielectron correlation effects, which require using the methods of multiparticle perturbation theory if we want to account for them.[26] Hence ordinarily the calculations of generalized polarizability constitute a complicated problem and require involved numerical methods (see Refs. 18 and 5). However, for hydrogenlike systems the solution can be obtained analytically.

To calculate the generalized dynamic polarizability of the hydrogen atom we use the method of the Coulomb Green's function in the coordinate representation. Other methods for calculating $\alpha(\omega,q)$ may also be employed (e.g., Sternheimer's method[27]). The use of Coulomb units allows extending the results to hydrogenlike ions.

The proposed method for calculating the generalized dynamic polarizability makes it possible to derive closed analytic expressions for the generalized polarizability of a state with an arbitrary principal quantum number. The idea of the method consists in representing the radial wave function R_{nl} of the hydrogen atom in terms of derivatives of the generating function of Laguerre polynomials:[28]

$$R_{nl}(r)=\frac{2^{l+1}}{n^{l+2}\sqrt{\Gamma(n-l)\Gamma(n+l+1)}} \times\frac{d^{n-l-1}}{dt^{n-l-1}}\left[\frac{r^le^{-\lambda r}}{(1-t)^{2l+2}}\right]\Bigg|_{t=0}, \tag{16}$$

where $\lambda=(1/n)(1+t)/(1-t)$.

Such a representation of the wave functions is convenient for calculating the matrix elements and hence the polarizability (7), since the operation of finding the derivatives is done after calculating the radial integrals in the matrix elements, which simplifies the entire calculation significantly. Actually one must calculate the matrix elements in which the radial wave functions of the hydrogen atom are replaced by the product $r^le^{-\lambda r}$. The final result for different states is then obtained by finding the derivatives of different orders of the expression obtained. This method is especially convenient if differentiation with respect to a parameter can be done symbolically (in our calculations we used the Mathematica 2.2 package of Wolfram Research, Inc.).

Let us show how the sum of matrix elements comprising the polarizability can be expressed in terms of a set of hypergeometric functions. Using the dispersion representation of the Green's function, we write the polarizability as the sum of two terms,:

$$\alpha(\omega,q)=X(E_0+\omega,q)+X(E_0-\omega,q), \tag{17}$$

where

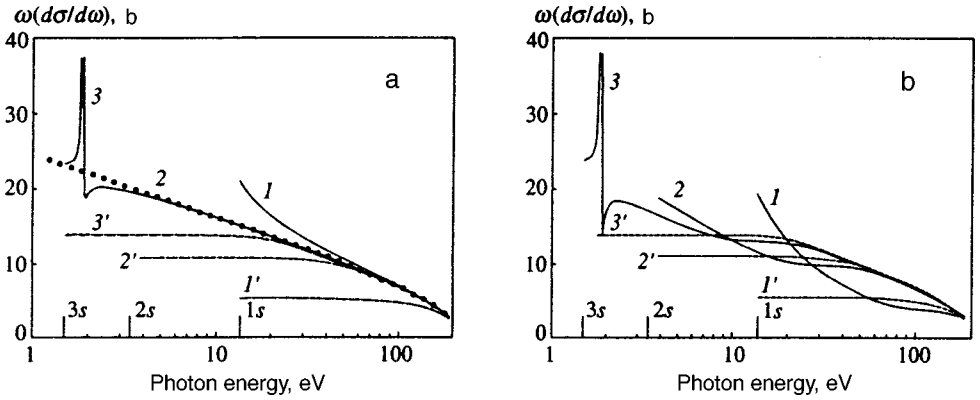


FIG. 1. Spectral curves of the bremsstrahlung cross sections (the Born approximation) for 220 eV electrons (a) and positrons (b) scattered by the hydrogen atom. The solid curves *1*, *2*, and *3* represent the total bremsstrahlung cross sections for the case of hydrogen atom in the $1s$-, $2s$-, and $3s$-states, respectively. The dashed curves *1′*, *2′*, and *3′* represent the ordinary bremsstrahlung cross sections for the hydrogen atom in the $1s$-, $2s$-, and $3s$-states, respectively. The dots correspond to the bremsstrahlung cross sections for an electron scattered by a point Coulomb potential (see the explanation in text).

$$X(E,q)=-\frac{i}{q}\langle 0|\exp(-i\mathbf{q}\cdot\mathbf{r}_1)G(\mathbf{r}_1,\mathbf{r}_2;E)\mathbf{r}_2\cdot\hat{\mathbf{q}}|0\rangle.$$

Here we have used the notation $\hat{\mathbf{a}}\equiv\mathbf{a}/|\mathbf{a}|$, where $\mathbf{a}$ is an arbitrary vector. Next we write the spherical-wave expansion of the Green's function:

$$G(\mathbf{r}_1,\mathbf{r}_2;E)=\sum_{l,m} G_l(r_1,r_2;E)Y_{lm}(\hat{\mathbf{r}}_1)Y^*_{lm}(\hat{\mathbf{r}}_2), \tag{18}$$

where for the Green's function of the radial Schrödinger equation we use the representation[29]

$$G_l(r_1,r_2;E)=\frac{2i}{\sqrt{r_1r_2}}(-1)^{l+1}\int_1^\infty\frac{d\xi}{\sqrt{\xi^2-1}}\left(\frac{\xi+1}{\xi-1}\right)^{i\nu}\times J_{2l+1}(2k\sqrt{r_1r_2}\,\sqrt{\xi^2-1}\,)\times\exp[ik\xi(r_1+r_2)], \tag{19}$$

where $k=\sqrt{2E}$, and $\nu=1/k$. Formally, this representation is valid for $E>0$. However, it can be analytically continued into the region where $E<0$. We do this analytic continuation in the formulas by expressing $X(E,q)$ in terms of hypergeometric functions.

Let us considered in detail the case where $l=0$, which means we have chosen the *s*-states of the hydrogen atom in the initial and final states of the process. Then, after integration over the angular variables has been performed, we can write the matrix element $X(E,q)$ as

$$X(E,q)=\frac{4}{n^5\Gamma^2(n)}\hat{D}_1\hat{D}_2M. \tag{20}$$

Here the differentiation operators have been defined as

$$\hat{D}_m f(t_m)=\frac{d^{n-1}}{dt_m^{n-1}}\left[\frac{f(t_m)}{(1-t_m)^2}\right]\Bigg|_{t_m=0},\quad m=1,2, \tag{21}$$

and the radial integral, after integration with respect to r_1 and r_2 and a change of variable, can be written as

$$M=2^9ik^3\int_0^1\frac{d\tau(1-\tau)\tau^{1-i\nu}(by-ax\tau)}{[(by-ax\tau)^2+q^2(y-x\tau)^2]^3}-2^9ik^5\int_0^1\frac{d\tau\,\tau^{2-i\nu}[5(by-ax\tau)^2-q^2(y-x\tau)^2]}{[(by-ax\tau)^2+q^2(y-x\tau)^2]^4}, \tag{22}$$

where $a=\lambda_1+ik$, $b=\lambda_1-ik$, $x=\lambda_2+ik$, $y=\lambda_2-ik$, and $\lambda_{1,2}=(1/n)(1+t_{1,2})/(1-t_{1,2})$.

The final expression for the integral (22) can be written as a sum of hypergeometric functions by expanding the integrands in partial fractions. The result, however, is extremely cumbersome, so that its explicit form is given in the Appendix.

3. RESULTS OF CALCULATIONS OF THE BREMSSTRAHLUNG CROSS SECTION FOR THE HYDROGEN ATOM AS TARGET

Figure 1 depicts the results of calculations of the cross section of bremsstrahlung formed in the process of the scattering of 220 eV electrons (Fig. 1a) and positrons (Fig. 1b) by the hydrogen atom. The calculations were done in the Born approximation. The solid curves *1*, *2*, and *3* represent the total bremsstrahlung cross sections and the dashed curves *1′*, *2′*, and *3′*, the ordinary bremsstrahlung cross sections for the hydrogen atom in the $1s$-, $2s$-, and $3s$-states, respectively. The small vertical lines indicate the values of the ionization potentials of the atomic shells.

Figures 1a and 1b show that the polarization mechanism plays an important role in forming the total bremsstrahlung spectrum for the hydrogen atom in the ground state and in an excited state. For photon frequencies close to the ionization potential of the hydrogen atom, the total bremsstrahlung cross section greatly exceeds the ordinary bremsstrahlung cross section. As the photon frequency increases, the total bremsstrahlung cross section monotonically decreases, and in absolute value remains larger than the ordinary bremsstrahlung cross section for the electron, while for the positron it becomes smaller than the ordinary bremsstrahlung

cross section. Such behavior of the total bremsstrahlung cross section (15) is due primarily to the constructive (for electrons) and the destructive (for positrons) contribution of the interference of the amplitudes of ordinary and polarization bremsstrahlung, $\omega(d\sigma/d\omega)_{\rm int}$. The effect has been described earlier. As shown in Refs. 12 and 30, the total bremsstrahlung spectrum in the photon frequency range $\omega \gg I_{1s}$ is effectively formed due to the braking of an incident electron in the Coulomb field of the bare atomic nucleus (the ''descreening'' effect).

The dots in Fig. 1a represent the bremsstrahlung spectrum for an electron scattered by a Coulomb potential with a charge $Z=1$. We see that when the photon energies are higher than the respective ionization potentials, the results of calculations of the total bremsstrahlung cross section almost fully agree with those calculated for the Coulomb potential.

A comparison of the spectral curves of the cross section of bremsstrahlung formed in the scattering of electrons and positrons by the hydrogen atom shows that for equal collision rates the contribution of the polarization bremsstrahlung mechanism increases with the principal quantum number. This effect is due to the increase in the radius of the orbit and hence to the weakening of the screening of the nucleus as the principal quantum number increases. Indeed, when bremsstrahlung is formed by the ordinary mechanism, small impact parameters are important, i.e., $r<R_{\rm at}$ (which means that the incident particles passes near the nucleus of the target atom). Hence an increase in the radius of the target atom leads to an increase in the ordinary bremsstrahlung cross section. On the other hand, polarization bremsstrahlung is formed most effectively at large impact parameters,[14,31] i.e., when the incident particle strongly polarizes the target atom.

Note that the spectral dependence of the total bremsstrahlung cross section for the case where an electron or positron is scattered by a hydrogen atom in the $3s$-state contains a narrow peak above the ionization potential of the $3s$-shell. The presence of such a maximum is due to a pole singularity in the polarizability in the photon frequency range corresponding to the $3s \to 2p$ transition.

Let us examine in detail the singularities exhibited in the generalized dynamic polarizability of excited states of the hydrogen atom. To this end we analyze Eq. (7), naming, for the sake of brevity, the first term in the braces A and the second term B.

For intermediate states $|n\rangle$ in the discrete spectrum we have $E_n<0$. If E_n is higher than the energy of the initial and final states E_0, i.e., $\omega_{n0}>0$, the term B behaves monotonically as the frequency ω increases, while the term A increases in a resonant manner near the poles $\omega \to \omega_{n0}$. The poles in the polarizability correspond to a real process, in which the hydrogen atom interacting with the incident particle is excited to the state $|n\rangle$ and then returns to its initial state $|0\rangle$, emitting a photon in the process. The divergence of the polarizability is removed by introducing an imaginary term $i\Gamma$ into the energy denominator, with Γ the width of the level. In the present paper the only frequencies ω we consider are those for which $\omega-\omega_0 \gg \Gamma$. Since the radiation width Γ is small, this condition holds up to frequencies that are extremely close to the polarizability poles.

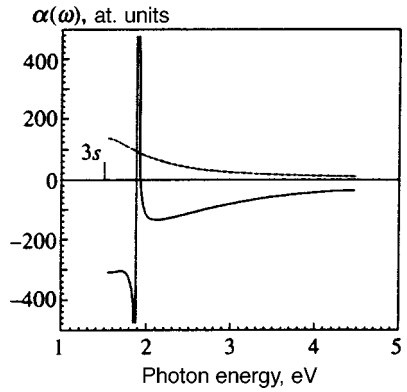


FIG. 2. The real (solid curve) and imaginary (dashed curve) parts of the dynamic dipole polarizability $\alpha(\omega)$ of the hydrogen atom in the $3s$-state near the polarizability pole $\omega=E_{3s}-E_{2p}=1.89$ eV.

If the initial state $|0\rangle$ of the hydrogen atom is not the ground state, then for some intermediate states $|n\rangle$ we have $0>E_0>E_n$, i.e., $\omega_{n0}<0$. If this condition is met, A behaves monotonically as the frequency ω increases, while B resonantly increases as $\omega \to -\omega_{n0}$. This divergence also corresponds to a real process in which the hydrogen atom, initially in the excited state $|0\rangle$, is de-excited and emits a photon and then, interacting with the incident particle, returns to the initial excited state. Here the excited state of the hydrogen atom must meet the condition $I \equiv -E_0 < -\omega_{n0}$, i.e., the polarizability poles $\omega=-\omega_{n0}$ are above the ionization threshold I of the given excited state $|0\rangle$.

Thus, the frequency dependence of polarizability contains poles corresponding to real processes of excitation and de-excitation of the atom in an intermediate state. Here the poles corresponding to de-excitation processes are above the ionization potential of the corresponding excited states. Hence the spectrum of polarization bremsstrahlung generated in collisions with the excited hydrogen atom must contain narrow lines not only below the ionization potential of the excited state but also above the ionization potential of that state. The only exception is the $2s$-state, since dipole transitions with de-excitation are forbidden for this state by selection rules.

Figure 2 depicts the spectral dependence of the dynamic polarizability $\alpha(\omega)$ of a hydrogen atom that is in its $3s$-state near the frequency corresponding to the $3s \to 2p$ transition. We see that the real part of the polarizability has a distinct pole, while the imaginary part is described solely by the term A and decreases monotonically with increasing frequency ω. For instance, the presence of a pole in the generalized polarizability above the ionization potential of the $3s$-shell leads to a narrow line in the bremsstrahlung spectrum (see Fig. 1).

Note that the interference term in the total cross section of the process is proportional to the real part of the polarizability, which changes sign as the pole is passed. The change in sign of the interference of the ordinary and polarization bremsstrahlung mechanisms implies that below the polarizability pole the bremsstrahlung cross section for a positron is larger than that for an electron, while above the pole the situation is just the opposite.

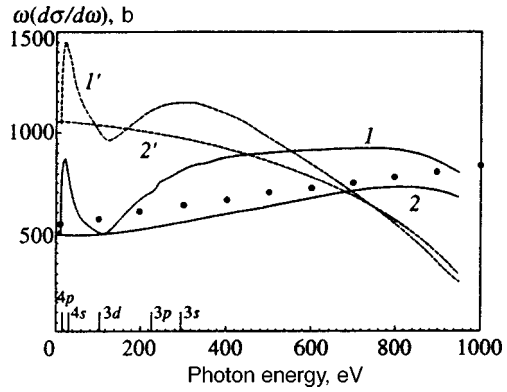


FIG. 3. Bremsstrahlung cross sections for 1 keV electrons scattered by the krypton atom. Curve *1* represents the total cross section (DPWA), curve *2* represents the ordinary bremsstrahlung cross section (DPWA), curve *1′* represents the total cross section (the Born approximation), and curve *2′* represents the ordinary bremsstrahlung cross section (the Born approximation). The dots stand for the data on the ordinary bremsstrahlung cross sections taken from Refs. 32 and 33.

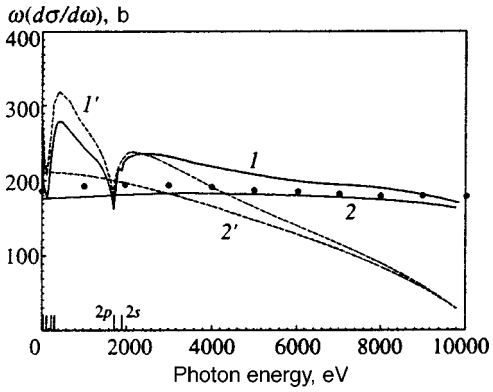


FIG. 5. Bremsstrahlung cross sections for 10 keV electrons scattered by the krypton atom. Curve *1* represents the total cross section (DPWA), curve *2* represents the ordinary bremsstrahlung cross section (DPWA), curve *1* represents the total cross section (the Born approximation), and curve *2* represents the ordinary bremsstrahlung cross section (the Born approximation). The dots stand for the data on the ordinary bremsstrahlung cross sections taken from Refs. 32 and 33.

4. RESULTS OF CALCULATIONS OF THE BREMSSTRAHLUNG CROSS SECTION WITH THE KRYPTON ATOM AS TARGET

The results of calculations of the cross sections of the total (solid curves *1*) and ordinary (solid curves *2*) bremsstrahlung formed in the scattering of 1-keV, 5-keV, and 10-keV electrons by the krypton atom are depicted in Figs. 3–5. The calculations were done in the DPWA approximation by Eq. (14). The small vertical lines indicate the values of the Hartree–Fock ionization potentials of the atomic subshells.

These calculations show that the polarization bremsstrahlung mechanism plays an important role in the formation of the total bremsstrahlung spectrum. Instead of a smooth curve characteristic of the spectral distribution of ordinary bremsstrahlung, the total bremsstrahlung cross section exhibits a complicated dependence on the photon frequency and has broad, high maxima and narrow dips near the ionization potentials. This pattern of the spectral dependence of the total bremsstrahlung cross section is determined both by the polarization component of the cross section, $\omega(d\sigma/d\omega)_{\rm pol}$ and by the interference term $\omega(d\sigma/d\omega)_{\rm int}$.

In the photon frequency range $\omega=10$–100 eV, the total bremsstrahlung spectrum for Kr exhibits a maximum due to the contribution to the amplitude $f_{\rm pol}$ of virtual dipole excitations of electrons from the outer $4s$- and $4p$-subshells. Such excitations are of a collective nature[26] and are characterized by a strong intershell correlation interaction of electrons from the ns- and np-subshells.[34] Amus'ya *et al.*[34] carried out a numerical calculation of the cross section of photoabsorption of the outer ns-subshells of atoms of inert gases using the random phase approximation with exchange (RPAE) and found that the photoabsorption cross section is strongly influenced by multielectron correlation effects. The expression (13) for the total bremsstrahlung cross section contains only one characteristic, $\alpha(\omega,q)$, which determines the dynamic response of the target to the external field of the incident particle. Atomic polarizability is closely linked to the photoabsorption cross section, so that to calculate $\alpha(\omega,q)$ of Kr we used the RPAE method with allowance for dipole transitions from all the atomic subshells.

In the photon frequency range $\omega=100$–1000 eV, the total bremsstrahlung spectrum for Kr acquires a second maximum, which is related to the dipole excitations into the continuous spectrum of electrons from the intermediate $3s$-, $3p$-, and $3d$-subshells of the krypton atom. Near the ionization potentials of these subshells, the multielectron correlation effects strongly influence the value of $\alpha(\omega,q)$. We used the RPAE method to account for these effects.

For photon frequencies exceeding 2000 eV, the total bremsstrahlung cross section is a monotonic function of ω, except for regions near the ionization potentials of the inner $1s$-, $2s$-, and $2p$-subshells of Kr. Below we analyze such behavior of the spectral dependence of bremsstrahlung.

The dashed curves *1* and *2* in Figs. 3–5 represents the results obtained in the first Born approximation. We see that the Born approximation gives a poor description of the total bremsstrahlung spectrum for all ω with the exception of the

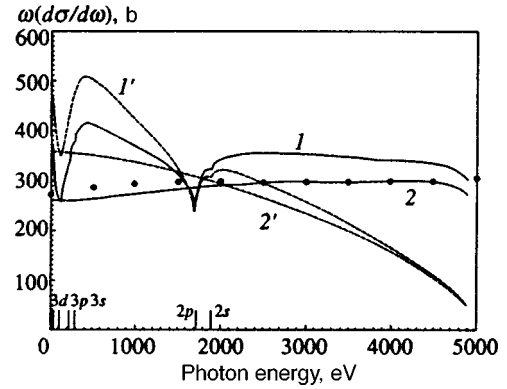


FIG. 4. Bremsstrahlung cross sections for 5 keV electrons scattered by the krypton atom. Curve *1* represents the total cross section (DPWA), curve *2* represents the ordinary bremsstrahlung cross section (DPWA), curve *1* represents the total cross section (the Born approximation), and curve *2* represents the ordinary bremsstrahlung cross section (the Born approximation). The dots stand for the data on the ordinary bremsstrahlung cross sections taken from Refs. 32 and 33.

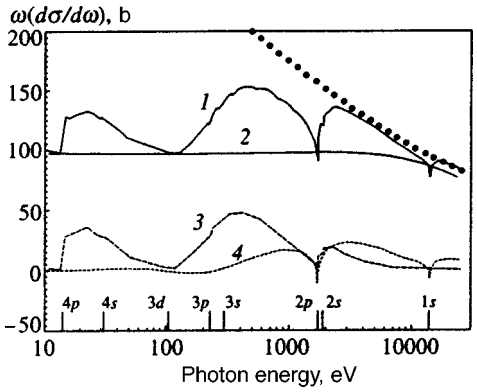


FIG. 6. The cross sections of total (curve *1*), ordinary (curve *2*), and polarization (curve *3*) bremsstrahlung for 25 keV atoms scattered by the krypton atom. Curve *4* represents the interference component of the total cross section. The dots correspond to the bremsstrahlung cross section for an electron scattered by a point Coulomb potential (see the explanation in text).

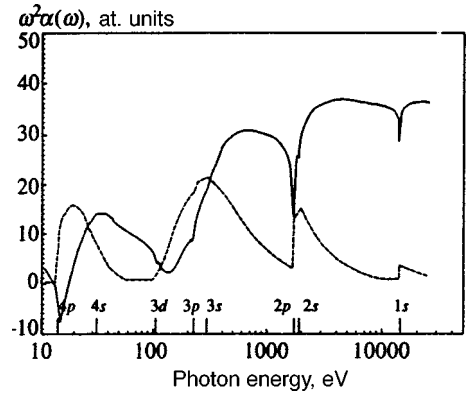


FIG. 7. The real (solid curve) and imaginary (dashed curve) parts of the dynamic dipole polarizability $\omega^2\alpha(\omega)$ of the krypton atom in the RPAE approximation.

low-frequency range, $\omega/\varepsilon_1 \leqslant 0.2$ (see also Ref. 35).

The discrepancy between the total bremsstrahlung cross sections obtained in the Born approximation and in the DPWA approximation, which are especially evident in Fig. 3, is related primarily to the difference in the ordinary bremsstrahlung cross sections (curve *2*). As noted earlier, ordinary bremsstrahlung is formed primarily at small distances, $r \leqslant R_{\rm at}$, where the distorting effect of the atomic potential causes the wave function of the incident electron to depart substantially from a plane wave. At the same time, polarization bremsstrahlung is formed at distances much larger than the atomic,[14,31] so that it is less sensitive to the approximations describing the motion of the incident electron.

The dots in Fig. 3–5 represent the results of calculations of the ordinary bremsstrahlung cross sections done by the relativistic DPWA approximation by Pratt *et al.*[32] and Seltzer and Berger.[33] The 5–10% discrepancy between their data and the curves *2* may arise because we ignored relativistic effects and higher-order multipole moments of the photon in our calculations and to the accuracy of the data of Refs. 32 and 33, which is estimated at 10 % by the researchers.

To study in greater detail the main tendencies in the behavior of the total bremsstrahlung cross section we examined the process of scattering of 25 keV electrons by the Kr atom. Curve *1* in Fig. 6 represents the total bremsstrahlung cross section and curve *2*, the ordinary bremsstrahlung cross section.

In the photon frequency range $\omega = 10$–1000 eV, the total bremsstrahlung cross section has two maxima. Within this spectral range, the total bremsstrahlung cross section (see Eq. (15)) is determined by the sum of the ordinary and polarization bremsstrahlung cross sections (curve *3*), while the contribution of the interference term $\omega(d\sigma/d\omega)_{\rm int}$ (curve *4*) is small.

Near the ionization potentials of the inner $1s$-, $2s$-, and $2p$-subshells the spectral distribution of the total bremsstrahlung cross section for the Kr atom as target exhibits deep narrow dips (cusps). The value of the total bremsstrahlung cross section at the tip of a cusp is smaller than that of the ordinary bremsstrahlung cross section. Similar singularities manifest themselves in the spectral dependence of the total bremsstrahlung cross section for the Ne and Ar atoms near the ionization potentials of the $1s$-subshell.[9]

Such singularities in the spectral dependence of the total bremsstrahlung cross section reflect the behavior of the real part of the generalized atomic dynamic polarizability near the ionization potentials of the inner atom subshells.

The amplitude (6) of polarization bremsstrahlung can be expressed in terms of the generalized atomic polarizability. The behavior of $\alpha(\omega,Q)$ as a function of ω proves to be similar for all values of Q. Hence qualitatively the ω-dependence of the polarization and interference terms in (15) can be estimated as follows:

$$\omega\left(\frac{d\sigma}{d\omega}\right)_{\rm pol} \sim \omega^4[(\mathrm{Re}\,\alpha(\omega))^2+(\mathrm{Im}\,\alpha(\omega))^2]g_1(\omega),$$

$$\omega\left(\frac{d\sigma}{d\omega}\right)_{\rm int} \sim -\omega^2\,\mathrm{Re}\,\alpha(\omega)g_2(\omega),$$

where $g_1(\omega)$ and $g_2(\omega)$ are smooth, monotonically decreasing functions of ω. This estimate shows that the interference term may be either positive or negative, depending on the sign of $\mathrm{Re}\,\alpha(\omega)$.

Figure 7 depicts the frequency dependence of the dynamic dipole polarizability in the spectral region $\omega = 10$ eV–25 keV. The solid curve depicts $-\omega^2\,\mathrm{Re}\,\alpha(\omega)$ and the dashed curve, $\omega^2\,\mathrm{Im}\,\alpha(\omega)$. For the present paper the atomic dynamic polarizability $\alpha(\omega)$ was calculated in the RPAE approximation with allowance for dipole transitions from all atomic subshells.

The imaginary part of the atomic polarizability of Kr has discontinuities near the ionization near the ionization potentials of the inner $1s$-, $2s$-, and $2p$-subshells (see Fig. 7). These discontinuities lead to deep dips (cusps) in $\mathrm{Re}\,\alpha(\omega)$, which in turn lead to deep, narrow dips in the total bremsstrahlung cross section. The occurrence of minima is determined by the fact that the interference part of the cross section, $\omega(d\sigma/d\omega)_{\rm int}$, in the frequency range under investigation is negative.

The origin of such singularities in the total bremsstrahlung cross section can be understood from the following rea-

soning. The imaginary part of the dipole polarizability is related to the photoabsorption cross section as follows: $\sigma_\gamma(\omega)=4\pi\omega\,\mathrm{Im}\,\alpha(\omega)/c$. Usually the absorption cross section for the inner atomic subshells is hydrogenlike. This means that at the photon frequency $\omega=I_{\mathrm{inner}}$ the cross section $\sigma_\gamma(\omega)$ has a discontinuity (see, e.g., the tables given in Ref. 36). Since the real and imaginary parts of the dynamic susceptibility are linked by a dispersion relation, the discontinuity in $\sigma_\gamma(\omega)$ leads to a cusp in $\mathrm{Re}\,\alpha(\omega)$. These discontinuities and dips manifest themselves in the form of singularities in the spectral dependence of the total bremsstrahlung cross section at frequencies $\omega\approx I_{\mathrm{inner}}$. The type of singularities in $\omega(d\sigma/d\omega)_{\mathrm{tot}}$ depends on the relationships between $\mathrm{Re}\,\alpha(\omega)$ and $\mathrm{Im}\,\alpha(\omega)$ and between $\omega(d\sigma/d\omega)_{\mathrm{int}}$ and $\omega(d\sigma/d\omega)_{\mathrm{pol}}$ and on the sign of $\omega(d\sigma/d\omega)_{\mathrm{int}}$. Generally, the total bremsstrahlung cross section may contain both dips and peaks.

In conclusion we discuss the bremsstrahlung process for photons whose energies exceed the ionization potential of the $1s$-subshell. Figure 6 shows that for $\omega>I_{1s}$ the total bremsstrahlung cross section decreases with increasing photon energy but exceeds the ordinary bremsstrahlung cross section. Such behavior of the total bremsstrahlung cross section (15) is due primarily to the ''descreening'' effect,[12,30] which leads to a situation in which the total bremsstrahlung spectrum of an electron in the frequency range $\omega\gg I_{1s}$ is effectively formed by the braking of the incident electron in the Coulomb field of the bare atomic nucleus, while the contribution of the polarization part of cross section is negligible (see Fig. 6).

The dots in Fig. 6 represent the bremsstrahlung spectrum of an electron scattered by a Coulomb potential with the charge $Z=36$. We see that for photon energies $\omega>I_{1s}$ the results of calculations of the total bremsstrahlung cross section (curve *1*) practically coincide with the results of calculations of the bremsstrahlung cross section for scattering by a Coulomb potential.

The present work was made possible by grants from the Russian Fund for Fundamental Research (Project 96-02-17922-a) and the International Science and Technology Center (Project 076-95).

APPENDIX

Below we give an explicit expression for calculating the generalized dynamic polarizability of the hydrogen atom in a state with principal quantum number n and zero orbital angular momentum $l=0$.

The polarizability is expressed as the sum of two terms containing hypergeometric functions ${}_2F_1$. Note that the final expression can be transformed into a more compact one by reducing the number of hypergeometric functions. However, we believe that the given representation is more convenient for doing numerical calculations:

$$\alpha_{ns}(\omega,q)=X(E_n+\omega,q)+X(E_n-\omega,q), \tag{A1}$$

where

$$X(E,q)=\frac{2^6}{n^5\Gamma^2(n)}\hat{D}_1\hat{D}_2M. \tag{A2}$$

The differential operators $\hat{D}$ have been defined in (21), and M has the form

$$\begin{aligned}
M=&-3\frac{\chi^2\alpha}{y^5\gamma q^3k(2-i\nu)(3-i\nu)}{}_2F_1(1,2-i\nu,4-i\nu,z_1)\\
&+3\frac{\chi\alpha^2}{y^5\beta q^3k(2-i\nu)(3-i\nu)}{}_2F_1(1,2-i\nu,4-i\nu,z_2)\\
&+6i\frac{\chi^2\alpha}{y^6\gamma q^3(3-i\nu)}{}_2F_1(1,3-i\nu,4-i\nu,z_1)\\
&-6i\frac{\chi\alpha^2}{y^6\beta q^3(3-i\nu)}{}_2F_1(1,3-i\nu,4-i\nu,z_2)\\
&+2i\frac{\chi^2(\alpha+2iq)}{y^5\gamma^2q^3(2-i\nu)(3-i\nu)}{}_2F_1(2,2-i\nu,4-i\nu,z_1)\\
&-2i\frac{\alpha^2(\chi-2iq)}{y^5\beta^2q^3(2-i\nu)(3-i\nu)}{}_2F_1(2,2-i\nu,4-i\nu,z_2)\\
&+6\frac{\chi^2k(\alpha+iq)}{y^6\gamma^2q^3(3-i\nu)}{}_2F_1(2,3-i\nu,4-i\nu,z_1)\\
&-6\frac{\alpha^2k(\chi-iq)}{y^6\beta^2q^3(3-i\nu)}{}_2F_1(2,3-i\nu,4-i\nu,z_2)\\
&+8i\frac{\chi^2k}{y^5\gamma^3q^2(2-i\nu)(3-i\nu)}{}_2F_1(3,2-i\nu,4-i\nu,z_1)\\
&+8i\frac{\alpha^2k}{y^5\beta^3q^2(2-i\nu)(3-i\nu)}{}_2F_1(3,2-i\nu,4-i\nu,z_2)\\
&-4i\frac{\chi^2k^2(\alpha+4iq)}{y^6\gamma^3q^3(3-i\nu)}{}_2F_1(3,3-i\nu,4-i\nu,z_1)\\
&+4i\frac{\alpha^2k^2(\chi-4iq)}{y^6\beta^3q^3(3-i\nu)}{}_2F_1(3,3-i\nu,4-i\nu,z_2)\\
&-24i\frac{\chi^2k^3}{y^6\gamma^4q^2(3-i\nu)}{}_2F_1(4,3-i\nu,4-i\nu,z_1)\\
&-24i\frac{\alpha^2k^3}{y^6\beta^4q^2(3-i\nu)}{}_2F_1(4,3-i\nu,4-i\nu,z_2),
\end{aligned} \tag{A3}$$

where the following notation has been used:

$$\begin{aligned}
&k=\sqrt{2E}\,,\quad \nu=\frac{1}{k},\\
&z_1=\frac{\chi x}{\gamma y},\quad z_2=\frac{\alpha x}{\beta y},\\
&\alpha=\lambda_1+ik+iq,\quad \beta=\lambda_1-ik+iq,
\end{aligned} \tag{A4}$$

$$\chi=\lambda_2+ik-iq, \quad \gamma=\lambda_2-ik-iq,$$
$$x=\lambda_2+ik, \quad y=\lambda_2-ik,$$
$$\lambda_1=\frac{1}{n}\frac{1+t_1}{1-t_1}, \quad \lambda_2=\frac{1}{n}\frac{1+t_2}{1-t_2}.$$

*)E-mail: lyalin@gnl.usr.pu.ru

[1] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed., Pergamon Press, Oxford (1991).
[2] R. H. Pratt, in *Fundamental Processes in Energetic Atomic Collisions*, H. O. Lutz, J. S. Briggs, and H. Kleinpoppen (eds.), Plenum Press, New York (1984).
[3] M. Ya. Amus′ya, V. M. Buĭmistrov, B. A. Zon, V. N. Tsytovich *et al.*, *Polarization Bremsstrahlung of Particles and Atoms* [in Russian], Nauka, Moscow (1987).
[4] M. Ya. Amus'ya, *Bremsstrahlung* [in Russian], Energoatomizdat, Moscow (1990).
[5] A. V. Korol and A. V. Solov'yov, J. Phys. B **30**, 1105 (1997).
[6] V. M. Buĭmistrov and L. I. Trakhtenberg, Zh. Éksp. Teor. Fiz. **69**, 108 (1975) [Sov. Phys. JETP **42**, 54 (1975)].
[7] V. M. Buĭmistrov, Yu. A. Krotov, and L. I. Trakhtenberg, Zh. Éksp. Teor. Fiz. **79**, 808 (1980) [Sov. Phys. JETP **52**, 411 (1980)].
[8] A. Dubois and A. Maquet, Phys. Rev. A **40**, 4288 (1989).
[9] A. V. Korol, A. G. Lyalin, and A. V. Solov'yov, J. Phys. B **30**, L115 (1997).
[10] M. Ya. Amus'ya, A. S. Baltenkov, V. B. Gilerson, Pis'ma Zh. Tekh. Fiz. **3**, 1105 (1977) [Sov. Tech. Phys. Lett. **3**, 455 (1977)].
[11] M. Ya. Amusia, Comments At. Mol. Phys. **11**, 123 (1982).
[12] M. Ya. Amusia, N. B. Avdonina, L. V. Chernysheva, and M. Yu. Kuchiev, J. Phys. B **18**, L791 (1985).
[13] N. B. Avdonina, M. Ya. Amus'ya, M. Yu. Kuchiev, and L. V. Chernysheva, Zh. Tekh. Fiz. **56**(2), 246 (1986) [Sov. Phys. Tech. Phys. **31**, 150 (1986)].
[14] M. Ya. Amus'ya, T. M. Zimkina, and M. Yu. Kuchiev, Zh. Tekh. Fiz. **52**, 1424 (1982) [Sov. Phys. Tech. Phys. **27**, 866 (1982)].
[15] É. T. Verkhovtseva, E. V. Gnatchenko, B. A. Zon, A. A. Nakipelov, and A. A. Tkachenko, Zh. Éksp. Teor. Fiz. **98**, 797 (1990) [Sov. Phys. JETP **71**, 443 (1990)].
[16] M. Ya. Amusia and A. V. Korol, J. Phys. B **24**, 3251 (1991).
[17] M. Ya. Amusia, L. V. Chernysheva, and A. V. Korol, J. Phys. B **23**, 2889 (1990).
[18] A. V. Korol, A. G. Lyalin, and A. V. Solov'yov, J. Phys. B **28**, 4947 (1995).
[19] A. V. Korol, A. G. Lyalin, A. S. Shulakov, and A. V. Solov'yov, J. Phys. B **28**, L155 (1995).
[20] A. V. Korol, A. G. Lyalin, A. V. Solov'ev, and A. S. Shulakov, Zh. Éksp. Teor. Fiz. **109**, 1174 (1996) [JETP **82**, 631 (1996)].
[21] A. V. Korol, A. G. Lyalin, and A. V. Solov'yov, Phys. Rev. A **53** 2230 (1996).
[22] A. V. Korol, A. G. Lyalin, A. S. Shulakov, and A. V. Solov'yov, J. Electron Spectrosc. Relat. Phenom. **79**, 323 (1996).
[23] T. M. Zimkina, A. S. Shulakov, A. P. Braĭto, A. P. Stepanov, and V. A. Fomichev, Fiz. Tverd. Tela (Leningrad) **26**, 1981 (1984) [Sov. Phys. Solid State **26**, 1201 (1984)].
[24] A. A. Tkachenko, E. V. Gnatchenko, and É. T. Verkhovtseva, Opt. Spektrosk. **78**, 208 (1995) [Opt. Spectrosc. **78**, 183 (1995)].
[25] A. Verweyen, G. R. Guthöhrlein, E. Gerhard *et al.*, in *Abstract 17th Int. Conf. X-Ray and Inner-Shell Processes* (Hamburg, Germany) (1996), p. 170.
[26] M. Ya. Amus'ya, *The Atomic Photoelectric Effect* [in Russian], Nauka, Moscow (1987).
[27] V. L. Yakhontov and K. Jungmann, Z. Phys. D **38**(2), 141 (1996).
[28] V. A. Fock, *Fundamentals of Quantum Mechanics*, Mir Publishers, Moscow (1978).
[29] L. P. Rapoport, B. A. Zon, and N. L. Manakov, *The Theory of Multiphoton Processes in Atoms* [in Russian], Atomizdat, Moscow (1978).
[30] A. V. Korol, J. Phys. B **25**, L341 (1992).
[31] B. A. Zon, Zh. Éksp. Teor. Fiz. **77**, 44 (1979) [Sov. Phys. JETP **50**, 21 (1979)].
[32] R. H. Pratt, H. K. Tseng, C. M. Lee, L. Kissel, C. MacCallum, and M. Riley, At. Data Nucl. Data Tables **20**, 175 (1977); Erratum, *ibid.* **26**, 477 (1981).
[33] S. M. Seltzer and M. J. Berger, At. Data Nucl. Data Tables **35**, 346 (1986).
[34] M. Ya. Amusia, V. K. Ivanov, N. A. Cherepkov, and L. V. Chernysheva, Phys. Lett. A **40**, 361 (1972).
[35] B. A. Zon, Zh. Éksp. Teor. Fiz. **107**, 1176 (1995) [JETP **80**, 655 (1995)].
[36] B. L. Henke, E. M. Gullikson, and J. C. Davis, At. Data Nucl. Data Tables **54**, 2 (1993).

Translated by Eugene Yankovsky

Crossing of quasilevels in the open Jaynes–Cummings model

M. Z. Smirnov*)

Laser Center, St. Petersburg State Institute of Precision Mechanics and Optics, 197101 St. Petersburg, Russia

(Submitted 24 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 474–483 (August 1998)

The nonlinear dynamics of an open quantum system consisting of a ''dressed'' atom, i.e., an atom coupled with an external multifrequency electromagnetic field, and a single quantized mode of an electromagnetic field is studied. At different crossing points of the quasilevels of the dressed atom, the average number of photons in the quantized mode may either increase without limit with the passage of time or oscillate within finite limits. In the latter case a decrease of the number of photons is accompanied by regularization of their statistics, which may become sub-Poissonian. © *1998 American Institute of Physics.* [S1063-7761(98)00808-7]

1. INTRODUCTION

The Jaynes–Cummings model, a closed quantum system consisting of a two-level atom and a single quantized mode of an electromagnetic field,[1] is a basis for more elaborate physical models of quantum optics and laser theory. The solution of the quantum mechanical equations describing the nonlinear dynamics of the Jaynes–Cummings model can be obtained in a fairly simple form.[2,3] Nevertheless, this solution describes important qualitative effects that manifest themselves in lasers and other complicated systems: pulsations of atomic level populations, changes in the statistics of the photons in a quantized mode of an electromagnetic field, and generation of squeezed quantum states of a field.[3–5] On the other hand, the development of experimental techniques has made it possible to experimentally realize a quantum system that is close in structure to the Jaynes–Cummings model, a one-atom maser.[6,7] Earlier theoretical results have been corroborated. Among these is the unusual nature of the dynamics of level populations and the dipole moment of the atom. Periodic collapses and revivals constitute a characteristic feature of this dynamics.[8,9]

As a result of the achievements in experiments (described above) and the development of theoretical quantum optics, many papers have appeared in which the Jaynes–Cummings model is generalized to the case of a multilevel atom, several quantized modes of an electromagnetic field, and multiphoton interaction. Some results were obtained without resorting to the rotating-wave approximation.[10,11] The review of these studies can be found in Refs. 3 and 10, and some latest results are given in Refs. 11–17.

A promising generalization of the Jaynes–Cummings model is the open model, which, in addition to the atom and the quantized field, may incorporate the electromagnetic-field vacuum[17] or an external classical field.[18,19] In the latter case the problem arises of the interaction between a ''dressed'' atom, i.e., an atom coupled with an external classical field, and a single quantized mode of an electromagnetic field. The general solution of this problem for the case of a monochromatic classical field was obtained by Kazakov,[18] and the new features that emerge in the dynamics of an open system if the classical field consists of several spectral components was analyzed in Ref. 19. In the later case the quasi-energy approach has proved to be useful. Here is a brief discussion of this approach.

The energy of dressed atoms is not conserved, so that the gain, absorption, and resonance-fluorescence spectra of such atoms cannot be interpreted in terms of transitions between energy levels. However, if the classical field has an equidistant frequency spectrum, the concepts of quasi-energy states and levels can be brought into the picture.[20,21] Spectroscopically, i.e., in studies of linear gain (absorption) of light by a dressed atom and of resonance fluorescence, these quasi-energy levels play the same role as the energy levels of an isolated atom. The difference between the energy levels and the quasi-energy levels becomes important when one considers nonlinear effects, such as the saturation of gain and absorption. The emission or absorption of a photon by an isolated atom is always accompanied be changes in the population of the energy levels, while emission or absorption of a photon by a dressed atom may take place without any change in the population of the quasi-energy states corresponding to different quasi-energy levels.[19] Here a change in the energy of the atom is balanced by an exchange of energy with the classical field. In Ref. 19 this type of interaction was named ''elastic,'' in contrast to the ordinary ''inelastic'' interaction, in which the emission or absorption of a photon changes the state of the dressed atom. Elastic interaction may lead to an effective conversion of the energy of the classical field into the energy of photons in the quantized mode. Here the quantum state of the atom remains unchanged and the statistics of the photons in the quantized mode is Poissonian or super-Poissonian, i.e., $\Delta n^2 \geqslant \bar{n}$, where $\bar{n}$ is the average and Δn^2 the variance of the number of photons in the quantized mode.

The position of the quasi-energy levels of a dressed atom depends on the parameters of the classical field (the different approaches to defining the energy levels of a dressed atom, which lead to the same results, are examined in Refs. 20–23). If these parameters (the intensity of the classical field, in

particular) are chosen appropriately, the quasilevels may cross, and at the crossing points the elastic and inelastic interaction mechanisms may act simultaneously.[19,24]

In an earlier paper[19] we examined the case where the quasi-energy levels do not cross and the frequency interval between the levels is much larger than the Rabi frequency. In the equations we can then ignore the terms describing the quantum interference of quasi-energy states[20] and arrive at a solution in analytic form. In the present paper we study the dynamics of the Jaynes–Cummings model with a dressed atom and crossing of the quasilevels. Actually we consider only one of the possible cases where such crossing occurs: the classical field is assumed modulated in amplitude (a triharmonic field), its median frequency coincides with the transition frequency between the atomic levels, and the Rabi frequency is an integral multiple of the modulation frequency. We analyze the temporal variations of the average number of photons, of the variance of the number of photons in the quantized mode, and of the populations of the quasi-energy states. It is assumed that initially the atom and the quantized mode are statistically independent, with the atom being in one of its quasi-energy states and the mode in a vacuum or coherent quantum state.

The plan of the paper is as follows. Section 2 contains the derivation of the Heisenberg equations for the annihilation (creation) operators of a photon in a quantized mode and the transition between quasi-energy states. Section 3 discusses the results of an analytical solution of these equations in various cases, while Sec. 4 does the same via a numerical solution. Finally, Sec. 5 briefly discusses the main results and formulates the conclusions.

2. HEISENBERG EQUATIONS FOR AN OPEN MODEL

Let us take a two-level atom interacting with an amplitude-modulated classical field:

$$E(t)=\tilde{E}g(\omega' t)\exp(i\Omega t)+\text{c.c.}, \tag{1}$$

where $\tilde{E}$ is the complex-valued amplitude of the field, ω' is the modulation frequency, Ω is the optical carrier frequency, and

$$g(\tau)=\exp(i\tau)[1+2a\ \cos\ \tau] \tag{2}$$

is the modulation function. By a classical field we mean laser light whose intensity is so high that we can ignore shot noise produced by photons.[25,26] The optical carrier frequency Ω exceeds the modulation frequency ω' by many orders of magnitude. The real parameter a is the percentage modulation. Thus, a classical field contains a central component at the frequency $\Omega_0=\Omega+\omega'$ and two sideband components at frequencies $\Omega_0\pm 2\omega'$. In what follows we assume that the central component coincides with the transition frequency ω of the transition between the atomic levels: $\Omega_0=\Omega+\omega'=\omega$. If we use the dipole-interaction and rotating-wave approximations,[2] the Hamiltonian of the atom in the external field can be written as

$$\hat{H}_a=\hbar\omega\hat{b}^\dagger\hat{b}-\hbar\omega'\sigma g(\omega' t)\exp[i(\Omega t+\psi)]\hat{b}+\text{h.c.}, \tag{3}$$

where ω is the transition frequency between the atomic energy levels, $\sigma\equiv|\mu_{01}\tilde{E}/\hbar\omega'|$ is the dimensionless field amplitude (the Rabi frequency normalized to the modulation frequency), with μ_{01} the matrix element of the dipole moment of the transition; $\psi\equiv\arg[\mu_{01}\tilde{E}/\hbar\omega']$; $\hat{b}\equiv|0\rangle_{a\ a}\langle 1|$ is the transition operator between the lower $|0\rangle_a$ and upper $|1\rangle_a$ energy levels of the atom, and h.c. stands for an expression that is the Hermitian conjugate of the first two terms on the right-hand side of Eq. (3). The Schrödinger equation with the Hamiltonian (3) has two orthonormalized solutions, $|\theta_0\rangle_a$ and $|\theta_1\rangle_a$, known as quasi-energy states.[19–21]

The quasi-energy levels can be found by analyzing the temporal variations of the wave functions describing the quasi-energy states.[21] Thus, we can show that each atomic energy level splits into two equally spaced sequences of quasilevels. Near the lower atomic level there emerge quasilevels with energies $\hbar\omega'(-\lambda+2n)$ and $\hbar\omega'(\lambda-\delta+2m+1)$, while near the upper level the quasilevels that emerge have the energies $\hbar\omega+\hbar\omega'(\delta-\lambda-(2m+1))$ and $\hbar\omega+\hbar\omega'(\lambda-2n)$, where $m,n=0,1,2,\ldots$, $\lambda=\sigma$ is the characteristic number, and $\delta\equiv(\Omega-\omega)/\omega'=-1$ is the detuning of the optical carrier frequency from the transition frequency (the detuning is normalized to the modulation frequency). This implies that if the Rabi frequency $|\mu_{01}\tilde{E}/\hbar|$ is an integral multiple of the modulation frequency ω', i.e., $\sigma=l=1,2,\ldots$, the quasilevels belonging to different sequences cross. The index $l=1,2,\ldots$ can be interpreted as the order of the parametric resonance that emerges as a result of quasilevel crossing.[20,27]

To go over to the open Jaynes–Cummings model, we must introduce a quantized mode of an electromagnetic field into the system. In describing the interaction between the quantized mode and the atom we use the dipole-interaction and rotating-wave approximations. The Heisenberg equations for the transition operator between the quasi-energy states, $\hat{c}=|\theta_0\rangle_{a\ a}\langle\theta_1|$, and the annihilation operator $\hat{a}$ for a photon in the quantized mode were derived in Ref. 19. The interaction between a dressed atom and the quantized mode is of a resonant nature if the frequency ω_q of the quantized mode coincides with one of the transition frequencies between the quasi-energy levels, which can be written as

$$\delta_q\equiv\frac{\Omega-\omega_q}{\omega'}=2m+1,$$

where m is an integer. In this case we can use the second rotating-wave approximation, leaving only the principal terms on the right-hand sides of the Heisenberg equations. For a resonance characterized by integral values of the indices, $l=0,1,2,\ldots$ and $m=0,\pm 1,\pm 2,\ldots$, the equations become

$$\frac{d}{d\tau}\hat{c}=(\hat{I}-2\hat{c}^\dagger\hat{c})(\alpha\hat{a}_q+\gamma\hat{a}_q^\dagger)+2(\beta\hat{a}_q-\beta^*\hat{a}_q^\dagger)\hat{c}, \tag{4}$$

$$\frac{d}{d\tau}\hat{a}_q=\beta^*(\hat{I}-2\hat{c}^\dagger\hat{c})+\gamma\hat{c}^\dagger-\alpha^*\hat{c}, \tag{5}$$

where

$$\alpha=(-1)^{l+m+1}\frac{1}{2}\kappa_q^* J_{l+m+1}(2al), \tag{6}$$

$$\gamma=(-1)^{l-m-1}\frac{1}{2}\kappa_q J_{l-m-1}(2al), \tag{7}$$

$$\beta=\frac{1}{2}\kappa_q^* \text{ at } m=-1, \quad \text{and } \beta=0 \text{ at } m\neq 1, \tag{8}$$

$$\kappa_q=\kappa\exp(-i\psi), \quad \hat{a}_q=\hat{a}\exp(i\omega_q t),$$

with κ the atom–quantized-mode coupling constant. If we introduce an effective Hamiltonian

$$\hat{H}_{\text{eff}}=i[\alpha\hat{a}_q\hat{c}^\dagger-\alpha^*\hat{a}_q^\dagger\hat{c}+\gamma\hat{a}_q^\dagger\hat{c}^\dagger-\gamma^*\hat{a}_q\hat{c} +(\beta^*\hat{a}_q^\dagger-\beta\hat{a}_q)(\hat{I}-2\hat{c}^\dagger\hat{c})], \tag{9}$$

Eqs. (4) and (5) can be written in canonical form:

$$\frac{d}{d\tau}\hat{c}=-i[\hat{c},\hat{H}_{\text{eff}}], \quad \frac{d}{d\tau}\hat{a}_q=-i[\hat{a}_q,\hat{H}_{\text{eff}}]. \tag{10}$$

3. NONLINEAR DYNAMICS: ANALYTICAL SOLUTIONS

In Secs. 3 and 4 we use the solutions of the system of equations (4) and (5) (or (10)) to examine the temporal variations of the average number of photons in the quantized mode,

$$\overline{n}=\langle\hat{a}^\dagger(\tau)\hat{a}(\tau)\rangle=\langle\hat{a}_q^\dagger(\tau)\hat{a}_q(\tau)\rangle, \tag{11}$$

of the variance of the number of photons in the quantized mode,

$$\Delta n^2=\langle\hat{n}^2\rangle-(\overline{n})^2, \tag{12}$$

$$\langle\hat{n}^2\rangle=\langle\hat{a}^\dagger(\tau)\hat{a}(\tau)\hat{a}^\dagger(\tau)\hat{a}(\tau)\rangle=\langle\hat{a}_q^\dagger(\tau)\hat{a}_q(\tau)\hat{a}_q^\dagger(\tau)\hat{a}_q(\tau)\rangle,$$

and of the populations of the quasi-energy states $|\theta_1\rangle_a$ and $|\theta_0\rangle_a$:

$$N_1=\langle\hat{c}^\dagger(\tau)\hat{c}(\tau)\rangle, \quad N_2=\langle\hat{c}(\tau)\hat{c}^\dagger(\tau)\rangle=1-N_1 \tag{13}$$

(the angle brackets denote quantum mechanical averaging). We start with the particular cases for which analytical solutions can be obtained.

The first case is

$$\alpha=\gamma=0. \tag{14}$$

If this condition is met, the Hamiltonian (9) describes the elastic Jaynes–Cummings model discussed in Ref. 19.

The second case is

$$\beta=0, \quad |\alpha|\neq|\gamma|. \tag{15}$$

Here the equations of motion (4) and (5) can be formally reduced to the ordinary (inelastic) model. Indeed, in the case $|\alpha|>|\gamma|$ we can introduce the unitarily transformed annihilation operator

$$\hat{e}_1\equiv\frac{\alpha\hat{a}_q+\gamma\hat{a}_q^\dagger}{\sqrt{|\alpha|^2-|\gamma|^2}}, \quad [\hat{e}_1,\hat{e}_1^\dagger]=\hat{I}. \tag{16}$$

Equations (4) and (5) then lead to the following equations of motion for the operators $\hat{e}_1$ and $\hat{c}$:

$$\frac{d}{d\tau}\hat{c}=\kappa_1(\hat{I}-2\hat{c}^\dagger\hat{c})\hat{e}_1, \quad \frac{d}{d\tau}\hat{e}_1=-\kappa_1\hat{c}, \tag{17}$$

where $\kappa_1=\sqrt{|\alpha|^2-|\gamma|^2}$. The equations are equivalent to those of the ordinary Jaynes–Cummings model, the only difference being that instead of the transition operator $\hat{b}$ they contain the quasi-energy transition operator $\hat{c}$.

In the case $|\alpha|<|\gamma|$ we can perform a similar transformation by introducing the annihilation operator

$$\hat{e}_2\equiv\frac{\alpha^*\hat{a}_q^\dagger+\gamma^*\hat{a}_q}{\sqrt{|\gamma|^2-|\alpha|^2}}, \quad [\hat{e}_2,\hat{e}_2^\dagger]=\hat{I}.$$

The equations of motion for the operators $\hat{e}_2$ and $\hat{c}$ become

$$\frac{d}{d\tau}\hat{c}^\dagger=\kappa_2(\hat{I}-2\hat{c}\hat{c}^\dagger)\hat{e}_2, \quad \frac{d}{d\tau}\hat{e}_2=-\kappa_2\hat{c}^\dagger, \tag{18}$$

where $\kappa_2=-\sqrt{|\gamma|^2-|\alpha|^2}$. The equations are equivalent to those of the ordinary Jaynes–Cummings model in which $\hat{b}$ is replaced by $\hat{c}^\dagger$. Actually, the difference between (17) and (18) is that in the first the quasi-energy state $|\theta_0\rangle_a$ acts as the lower state of the atom and $|\theta_1\rangle_a$ as the upper, while in the second $|\theta_1\rangle_a$ is the lower state and $|\theta_0\rangle_a$ is the upper.

The third case is

$$\beta=0, \quad |\alpha|=|\gamma|. \tag{19}$$

Using Eqs. (4) and (5), we can show that the operator $\hat{A}=\alpha\hat{a}_q+\gamma\hat{a}_q^\dagger$ is a constant of motion of the system:

$$\hat{A}(\tau)=\hat{A}(0)=\alpha\hat{a}_0+\gamma\hat{a}_0^\dagger,$$

where $\hat{a}_0\equiv\hat{a}_q(0)$. Taking the derivatives with respect to τ of the right- and left-hand sides of Eq. (5), we get $(d^2/d\tau^2)\hat{a}_q=0$, which yields

$$\hat{a}_q(\tau)=\hat{a}_0+(\gamma\hat{c}_0^\dagger-\alpha^*\hat{c}_0)\tau, \tag{20}$$

where $\hat{c}_0\equiv\hat{c}(0)$.

Suppose that initially the atom was in one of its quasi-energy states and the quantized mode was in the coherent state $|\nu\rangle_f$ (Ref. 25):

$$\hat{a}_0|\nu\rangle_f=\nu|\nu\rangle_f. \tag{21}$$

Then, using the solution (20), we arrive at the following relationships for the average number of photons and the variance of the number of photons in the quantized mode:

$$\overline{n}=|\nu|^2+|\alpha|^2\tau^2, \tag{22}$$

$$\Delta n^2=|\nu|^2+[|\alpha|^2(2|\nu|^2+1)-\alpha^*\gamma(\nu^*)^2-\alpha\gamma^*\nu^2]\tau^2, \tag{23}$$

where $|\alpha|=|\gamma|$ Without loss of generality we can assume that the coupling constant κ_q is real. Then, according to (6) and (7), α and γ a re also real. In two cases the statistics of the photons in the quantized mode is Poissonian ($\Delta n^2=\overline{n}$): (a) α and γ are of the same sign and ν is real, and (b) α and γ have opposite signs and ν is purely imaginary. In all other cases the photon statistics is super-Poissonian, i.e., $\Delta n^2>\overline{n}$.

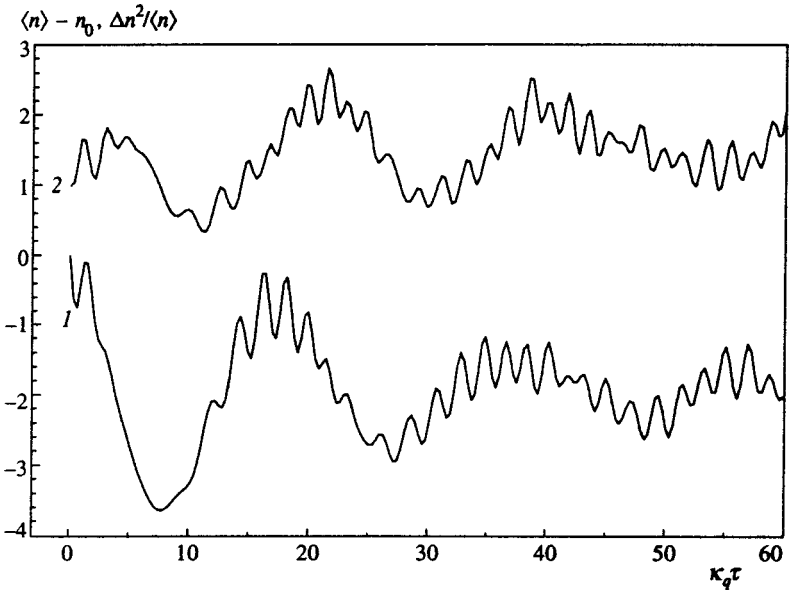


FIG. 1. Temporal variations of the average number of photons and of the variance of the number of photons in a quantized mode at $\alpha=\gamma=-0.29\kappa_q$ and $\beta=0.5\kappa_q$ (κ_q is real). This corresponds to the resonance $l=1$ and $m=-1$ at $a=0.968$ (a is the percentage modulation). At $\tau=0$ the mode is in the coherent quantum state $|\nu\rangle_f$ ($\nu=2.5$) and the dressed atom is in the quasi-energy state $|\theta_1\rangle_a$. Curve *1* represents the deviation of the average number $\langle n\rangle$ of photons from the initial value $n_0=|\nu|^2$ and curve *2* the relative variance $\Delta n^2/\langle n\rangle$ of the photon number.

4. NONLINEAR DYNAMICS: NUMERICAL ANALYSIS

Above we discussed three cases in which only one of the two possible types of interaction is present, elastic or inelastic. When both are present, the system of equations (4) and (5) can probably be solved only numerically. In the numerical solution we used the matrix representation of operators entering into Eqs. (4) and (5) in the base of quasi-energy states. Programs were written using the Mathematica 2.2 and Delphi 2.0 packages. The programs were tested and the accuracy of the calculations was checked by comparison with the analytical solutions of Sec. 3. Below we discuss some results of the numerical analysis of the dynamics of the open model.

Figures 1 and 2 depict the temporal variations of the average value and variance of the number of photons in the quantized mode (Fig. 1) and of the population of the quasi-energy state $|\theta_1\rangle$ (Fig. 2) for the open model at the point of quasilevel crossing determined by the indices $l=1$ and $m=-1$ at $a=0.968$ ($\alpha=\gamma=-0.29\kappa_q$ and $\beta=0.5\kappa_q$). The quantities $\langle n(\tau)\rangle-n_0$ (for curve *1*) and $\eta(\tau)=\Delta n^2(\tau)/\langle n(\tau)\rangle$ (for curve *2*) are laid off on the vertical axis in Fig. 1 (here $\langle n(\tau)\rangle\equiv\overline{n(\tau)}$ and $n_0=|\nu|^2$ are the average number of photons in the quantized mode at the current and initial moments in time, respectively). The photon statistics at time τ is sub-Poissonian if $\eta(\tau)<1$. Initially (at $\tau=0$) the atom was in the quasi-energy state $|\theta_1\rangle_a$ and the quantized mode was in the coherent state $|\nu\rangle_f$, where $\nu=2.5$. As in the ordinary Jaynes–Cummings model,[3] there are fast and slow oscillations. The average number of photons (curve *1* in Fig. 1) may exceed the initial number of photons by more than one, which is possible because of energy exchange with the classical field. Here the absorption of photons in the quantized mode is accompanied by regularization of their statistics, which becomes sub-Poissonian (curve *2* in Fig. 1). The squeezing factor $\xi(\tau)=\eta(\tau)-1$ reaches its minimum value $\xi\approx-0.6$. A possible interpretation of this goes as follows: there is a positive correlation between successive events of photon absorption from the

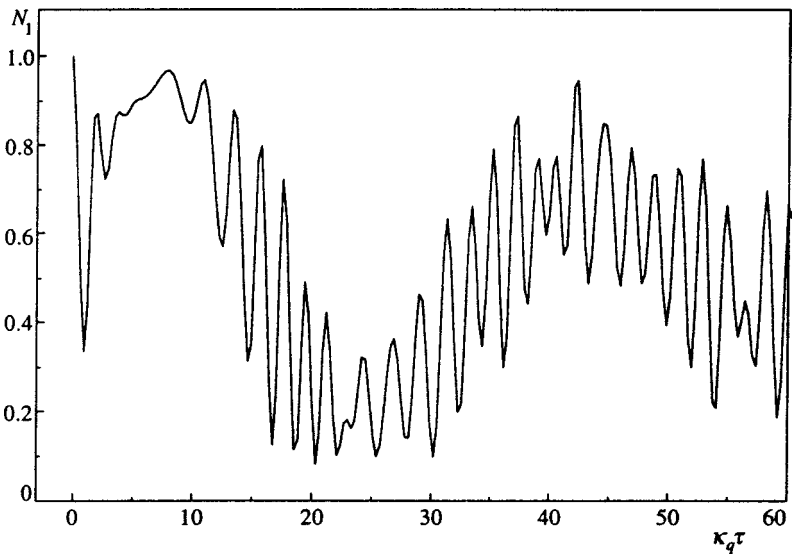


FIG. 2. Temporal variation of the average population N_1 of the quasi-energy state $|\theta_1\rangle_a$ for the same initial conditions and values of the parameters as in Fig. 1.

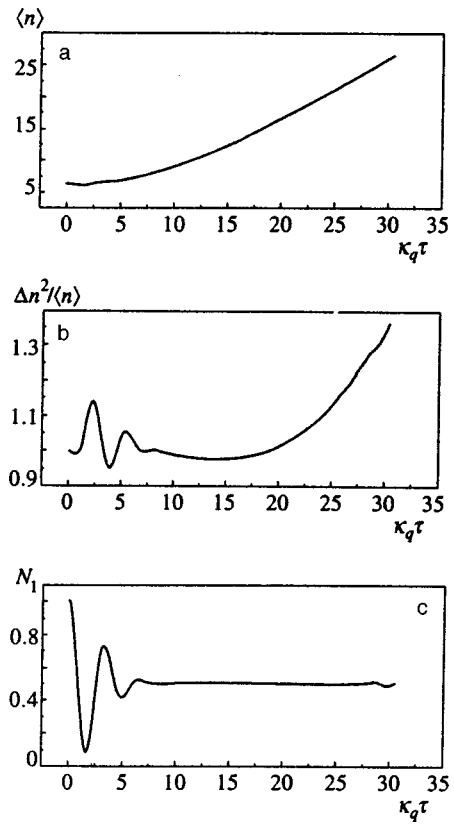


FIG. 3. Temporal variations of (a) the average number $\langle n\rangle$ of photons in the quantized mode, (b) the relative variance $\Delta n^2/\langle n\rangle$ of the number of photons in the quantized mode, and (c) the average population N_1 of the quasi-energy state $|\theta_1\rangle_a$, at $\alpha=0.116\kappa_q$, $\beta=0$, and $\gamma=0.256\kappa_q$. This corresponds to the resonance $l=1$ and $m=0$ at $a=0.75$ (a is the percentage modulation).

quantized mode (photons tend to be absorbed in pairs, triples, etc.), which leads to antigrouping and the sub-Poissonian photon statistics.[26] This is similar to the generation of squeezed states in two-photon absorption.[28] The reader will also note that there is no distinct correlation between the oscillations of the population of state $|\theta_1\rangle_a$ (Fig. 2) and those of the average number of photons in the quantized mode (curve *1* in Fig. 1). At different quasilevel crossing points the dynamics of the open model differs significantly. In particular, there can be a limitless increase in the number of photons in the quantized mode (as long as the classical-field fixed-amplitude approximation remains valid). An example of this kind is shown in Fig. 3, which illustrates the dynamics of (a) the average number of photons, (b) the relative variance of the number of photons, and (c) the population of the state $|\theta_1\rangle_a$, at $l=1$, $m=0$, and $a=0.75$ ($\alpha=0.11\kappa_q$, $\beta=0$, and $\gamma=0.256\kappa_q$). The initial conditions are the same as in Figs. 1 and 2.

5. CONCLUSION

We have studied the nonlinear dynamics of an open system consisting of a two-level atom interacting with a classical electromagnetic field and a quantized mode of an electromagnetic field. The parameters of the classical field are assumed fixed. It is also assumed that this field has an equidistant frequency spectrum. The interaction of both fields with the atom is described in dipole-interaction and rotating-wave approximations.

A two-level atom coupled with a classical field can be interpreted as a new quantum object, a dressed atom. The spectroscopic properties of a dressed atom are determined by the discrete quasi-energy levels, which shift as the classical field intensity changes (the dynamic Stark effect). If the natural frequency of the quantized mode coincides with the transition frequency between the quasi-energy levels, the interaction of the dressed atom and the quantized mode is of a resonant nature, and in describing such an interaction mathematically we can use the second rotating-wave approximation.

The difference between the energy levels of an isolated atom and the quasi-energy levels of a dressed atom manifests itself in the analysis of nonlinear effects (the effect of saturation of the population difference, in particular). In contrast to an isolated atom, the emission of absorption of a photon by a dressed atom may not be accompanied by a change in the populations of the quasi-energy states. Such an interaction, known as ''elastic,'' leads to a dramatic difference between the dynamics of an open model and that of the ordinary Jaynes–Cummings model. Elastic interaction increases the oscillation amplitude of the average number of photons in the quantized mode in relation to the ordinary Jaynes–Cummings model. Successive absorption (or emission) by the atom of several photons from (or into) the quantized mode becomes possible due to energy exchange with the classical field.

At a point where quasilevels cross, both interaction mechanisms (the elastic and the ordinary inelastic) may act simultaneously. However, in some cases only one mechanism is present, and for these cases an analytical solution of the Heisenberg equation of the open model can be obtained. For the general case a method of numerical investigation of the dynamics of the open model has been developed. One interpretation of the results of calculations is that the events in which a sequence of photons are absorbed are correlated with each other and lead to a ''smoothing-out'' of the photon noise in the quantized mode. After several photons are absorbed, the statistics of the remaining photons may become sub-Poissonian. At some quasilevel crossing points the dynamics of the open model is different: the average number of photons in the quantized mode increases without limit (as long as one can ignore the variations in the amplitude of the classical field) and their statistics remains Poissonian or becomes super-Poissonian.

Although all the numerical results refer to the particular case of quasi-energy level crossing in an amplitude-modulated classical field, the expression (9) for the effective Hamiltonian and the analytical solutions of Sec. 3 are more general and are independent of the spectrum of the classical field (the only requirement being that the quasi-energy levels cross). Actually, the characteristics of the classical field determine only the specific form of the coefficients α, β, and γ in the analytical expressions.

*)E-mail: lc@lacen.spb.su

[1] E. T. Jaynes and F. W. Cummings, Proc. IEEE **51**, 89 (1963).
[2] L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms*, Wiley, New York (1975).
[3] H.-I. Yoo and J. H. Eberly, Phys. Rep. **118**, 239 (1985).
[4] P. Meystre and M. S. Zubairy, Phys. Lett. A **89**, 390 (1982).
[5] F.-X. Zhao, M. Orszag, J. Bergou, and S.-J. Zhu, Phys. Lett. A **137**, 479 (1989).
[6] D. Meschede, H. Walther, and G. Müller, Phys. Rev. Lett. **54**, 551 (1985).
[7] G. Rempe, H. Walter, and N. Klein, Phys. Rev. Lett. **58**, 353 (1987).
[8] J. H. Eberly, N. B. Narozhny, and J. J. Sanchez-Mondragon, Phys. Rev. Lett. **44**, 1323 (1980).
[9] N. B. Narozhny, J. J. Sanchez-Mondragon, and J. H. Eberly, Phys. Rev. A **22**, 236 (1981).
[10] V. V. Dodonov, V. I. Man'ko, and S. M. Chumakov, in *Proc. (Trudy) of the P. N. Lebedev Physics Institute* [in Russian], Vol. **176, Moscow (1986), p.** 57.
[11] J. L. Gruver, J. Aliaga, H. A. Cerdeira, and A. N. Proto, Phys. Rev. A **50**, 5274 (1994).
[12] M. M. Ashraf, Phys. Rev. A **50**, 5116 (1994).
[13] H. Huang and H. Fan, Phys. Lett. A **166**, 308 (1992).
[14] H. Huang and H. Fan, Phys. Lett. A **159**, 323 (1991).
[15] A. Ya. Kazakov, Phys. Lett. A **206**, 229 (1995).
[16] A. Ya. Kazakov, Opt. Spektrosk. **81**, 549 (1996) [Opt. Spectrosc. **81**, 498 (1996)].
[17] H. T. Dung and A. S. Shumovski, Phys. Lett. A **169**, 379 (1992).
[18] A. Ya. Kazakov (private communication).
[19] M. Z. Smirnov, Zh. Éksp. Teor. Fiz. **112**, 818 (1997) [JETP **85**, 441 (1997)].
[20] M. Z. Smirnov, Phys. Rev. A **52**, 2195 (1995).
[21] M. Z. Smirnov, Kvant. Élektron. (Moscow) **22**, 903 (1995) [Quantum Electron. **25**, 871 (1995)].
[22] J. H. Shirley, Phys. Rev. B **138**, 979 (1965).
[23] C. Cohen-Tannoudji, in *Cargese Lectures in Physics*, Vol. 2, M. Levy (Ed.), Gordon and Breach, New York (1968), p. 347.
[24] M. Z. Smirnov, J. Opt. Soc. Am. B **9**, 2171 (1992).
[25] W. H. Louisell, *Quantum Statistical Properties of Radiation*, Wiley, New York (1973).
[26] D. F. Smirnov and A. S. Troshin, Usp. Fiz. Nauk **153**, 233 (1987) [Sov. Phys. Usp. **30**, 851 (1987)].
[27] B. Blind, P. R. Fontana, and P. Thomann, J. Phys. B **13**, 2717 (1980); P. Thomann, *ibid.* **9**, 2411 (1976); **13**, 1111 (1980).
[28] R. J. Horowicz, Europhys. Lett. **10**, 537 (1989).

Translated by Eugene Yankovsky

Coherent suppression of the EPR-nutation signal in quartz

V. S. Kuz'min*)

Institute of Solid-State and Semiconductor Physics, Academy of Sciences of Belarus, 220072 Minsk, Belarus
(Submitted 31 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 484–491 (August 1998)

Bloch equations are used to derive an analytic expression for the integral nutation-signal decay rate in a two-level quantum system. It is found that when inhomogeneous line broadening is taken into account, the temporal decay of the response is due not only to coherent suppression but also to additional decay whose rate depends on the ratio of the Rabi frequency to the inhomogeneous linewidth. The results are used to explain the anomalous (Rabi-frequency-dependent) EPR-nutation decay detected in quartz in the experiments of R. Boscaino, F. M. Gelardi, and J. P. Corb [Phys. Rev. B **48**, 7077 (1993)]. Contrary to the statement made by these researchers that this decay is non-Bloch, it is found that the process can be described by Bloch equations without any modification.
[S1063-7761(98)00908-1]

1. INTRODUCTION

The phenomenon of transient nutations observed in various parts of the electromagnetic spectrum reflects the temporal evolution of a quantum system that was in a state of thermodynamic equilibrium prior to excitation and develops into a new state in response to a resonant electromagnetic field. If the system is an ensemble of identical particles, nutation decay takes place at a dephasing rate T_2^{-1}. But if the ensemble driven by the field consists of particles with different transition frequencies, then even without irreversible relaxation, nutation decay occurs because of what is known as the coherent suppression effect. The essence of this effect is that the statistics of the spread of the values of the transition frequencies (inhomogeneous broadening) gives rise to a decay of coherent oscillations of the ensemble average of an observable (polarization, magnetization) of the quantum system.[1] (The other factors that lead to coherent suppression of nutation are the nonuniformity of the driving field over the volume of the sample and the spread in the orientations of the dipole moments of transitions.[1]) The limiting cases of broad and narrow inhomogeneously broadened lines are usually considered in treatments of nutation. In the first case, coherent suppression is described by a Bessel function and irreversible decay follows an exponential law. In the second, the decay is due only to irreversible relaxation. In the intermediate case, where the finiteness of the linewidth must be taken into account, the decay law is more complicated than in the above limits.

Despite the simplicity and generality of the nutation effect, there are relatively few works on the subject. Primarily this is due to the experimental difficulties associated with the discrimination of a particular decay mechanism. On the other hand, the problem of nutation decay is interesting from the viewpoint of verifying the validity of the Bloch equation in describing coherent phenomena in high-power driving fields. Recently this aspect has been examined by Boscaino *et al.*,[2] who studied EPR nutations of $[AlO_4]^0$ and E^1 centers in quartz. In the experiment described in Ref. 2 the actual source of coherent suppression was inhomogeneous broadening, which the researchers considered in the broad-line approximation. After extracting the oscillating part of the signal, described by a zeroth-order Bessel function, the researchers found that the signal envelope decays exponentially with a damping constant

$$\beta = \frac{1}{2T_2} + \alpha\chi, \tag{1}$$

where α is a constant coefficient and χ is the Rabi frequency.

According to Boscaino *et al.*,[2] this is an unexpected result, since after coherent suppression due to inhomogeneous broadening has been isolated the decay of the response in the broad-line approximation occurs solely because of irreversible relaxation. Such behavior of the signal is not described by Bloch equations, which prompted the researchers to conclude that the additional nutation decay dependent on the field amplitude is generally inherent in homogeneous systems. Recently, Fedoruk[3] has checked the above statement in NMR experiments on protons in glycerin. It was found that in a large range of field amplitudes nutation decay is due solely to irreversible relaxation and is independent of the field amplitude. Calculations of nutation decay in the high-field approximation of Torrey[4] also provided no satisfactory explanation of the observed anomaly.[5] The present paper is an attempt to study nutation decay in the general case without resorting to approximations. It is believed that the use of the broad-line approximation by Boscaino *et al.*[2] was unjustified, since in Ref. 2 the ratio χ/σ (σ is the inhomogeneous halfwidth of the line) reached 1.4. Under such conditions one must take into account the finiteness of the linewidth, which may lead to ''additional'' damping od the response. In other words, the goal was to establish whether under the conditions of the experiment described in Ref. 2 the broad-line approximation is adequate for calculating such a subtle effect as nutation decay.

2. THEORETICAL ANALYSIS

Let us examine the expression for the V-component of the magnetization of a spin system ($S=1/2$) with allowance for irreversible relaxation:[4]

$$V(t,\Delta)=V_0\,\frac{\chi\,\sin(t\sqrt{\Delta^2+\chi^2})}{\sqrt{\Delta^2+\chi^2}}\exp\left(-\frac{t}{2T_2}\right), \tag{2}$$

where V_0 is the equilibrium value of magnetization, Δ is the spread of the frequencies of the spin packets of an inhomogeneously broadened line, and $t>0$.

For further analysis of (2) we must average the above expression over the envelope of the inhomogeneously broadened line:

$$\langle V(t)\rangle=\int_{-\infty}^{\infty}V(t,\Delta)g(\Delta)d\Delta. \tag{3}$$

Assuming that the form factor $g(\Delta)$ is a Lorentzian, we can use the integral representation[6]

$$\frac{\sin(t\sqrt{\Delta^2+\chi^2})}{\sqrt{\Delta^2+\chi^2}}=\int_0^t J_0(\chi\sqrt{t^2-x^2})\cos\,\Delta x\,dx.$$

After repeated integration, Eq. (3) becomes

$$\langle V(t)\rangle=V_0\chi t\int_0^1 J_0(\chi t\sqrt{1-x^2})\exp\left[-t\left(\sigma x+\frac{1}{2T_2}\right)\right]dx, \tag{4}$$

where $J_0(z)$ is the zeroth-order Bessel function.

In the above expression the role of the finiteness of the linewidth in the signal decay clearly manifests itself. In order to clarify this role, the expression (4) was evaluated numerically for different χ/σ ratios without allowing for irreversible relaxation (Fig. 1). For the sake of comparison, Fig. 1 also depicts the nutation signal in the broad-line approximation of Ref. 4:

$$\langle V(t)\rangle=V_0\,\frac{\chi}{\sigma}\,J_0(\chi t). \tag{5}$$

We see that for $\chi/\sigma\leqslant 1$ the signals (4) and (5) coincide except for the initial section (Fig. 1a). However, for $\chi/\sigma>1$ some discrepancy appears, which grows with χ (Figs. 1b and 1c). The initial section resembles more and more a damped sinusoid, with the damping constant depending on χ/σ.

To estimate the decay we calculate the integral decay rate for signal (4):

$$\Gamma=\frac{\langle V(0)\rangle}{\int_0^\infty\langle V(t)\rangle dt}. \tag{6}$$

For $\langle V(0)\rangle$ we take the value of the signal at the time t_m of the initial nutation burst, $\langle V(t_m)\rangle$. The integral in the denominator of (6),

$$\langle\langle V(t)\rangle\rangle=\int_0^\infty\langle V(t)\rangle dt=V_0\chi\int_0^\infty t\,dt\int_0^1 J_0(\chi t\sqrt{1-x^2})\times\exp\left[-t\left(\frac{1}{2T_2}+\sigma x\right)\right]dx,$$

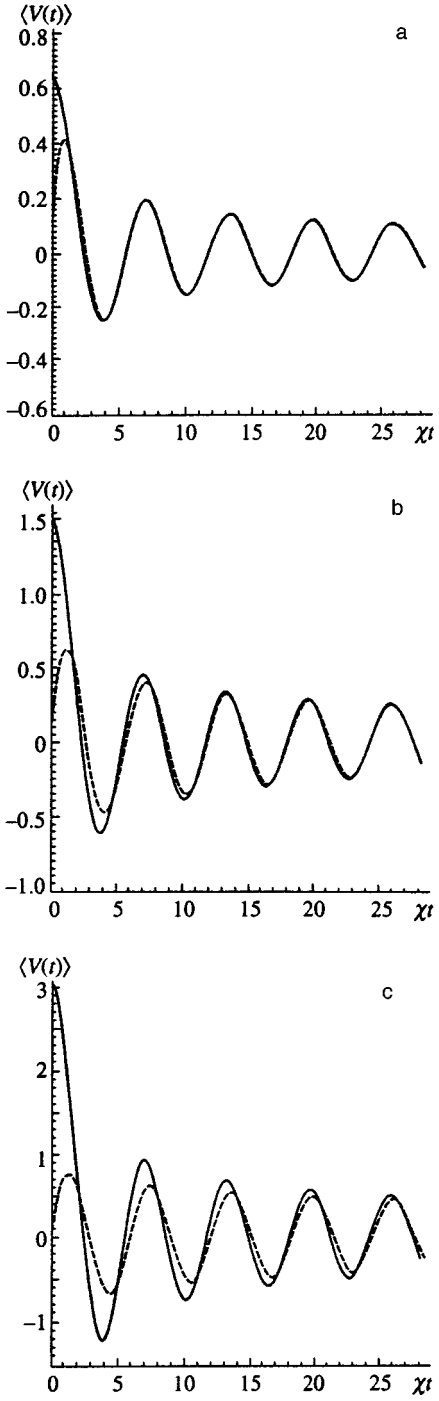


FIG. 1. Temporal dependence of the nutation signal at $\chi/\sigma=0.64$ (a), 1.5 (b), and 3.0 (c). The solid curves represent the signal described by Eq. (4); the dashed curves represent the signal described by formula (5).

after repeated integration becomes

$$\langle\langle V(t)\rangle\rangle=\chi\left[\sqrt{\chi^2+\frac{1}{4T_2^2}}\left(\sigma+\sqrt{\chi^2+\frac{1}{4T_2^2}}\right)\right]^{-1}. \tag{7}$$

Then the integral decay rate is

$$\Gamma=\langle V(t_m)\rangle\sqrt{\chi^2+\frac{1}{4T_2^2}}\left(\sigma+\sqrt{\chi^2+\frac{1}{4T_2^2}}\right)\chi^{-1}. \tag{8}$$

Let us analyze (8) for various field-amplitude values. At small amplitudes, $\langle V(t_m)\rangle$ varies approximately according

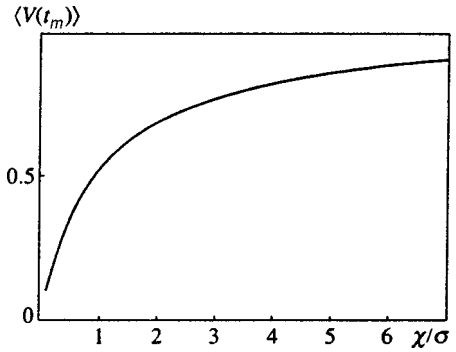


FIG. 2. Field dependence of the amplitude $\langle V(t_m)\rangle$ in the initial spike in the nutation.

linearly, $\langle V(t_m)\rangle \simeq \chi/\sigma$, while at large amplitude it asymptotically approaches unity (Fig. 2). Hence for moderate values of χ the integral decay rate is

$$\Gamma = \sqrt{\chi^2 + \frac{1}{4T_2^2}} + \frac{\chi^2 + 1/4T_2^2}{\sigma} = \left[\int_0^\infty J_0(\chi t)\exp\left(-\frac{t}{2T_2}\right)dt\right]^{-1} + \frac{\chi}{\sigma} \times \left[\int_0^\infty \sin\chi t\,\exp\left(-\frac{t}{2T_2}\right)dt\right]^{-1}, \quad (9)$$

i.e., the total integral decay rate is the sum of two rates, the first being a reflection of the signal decay in the broad-line approximation and the second, in the narrow-line approximation. For $\chi < \sigma$ the predominant decay is the one due to coherent suppression and irreversible relaxation, while the decay corresponding to the second term acts as a small addition, whose ''weight'' increases with χ. If $\chi > T_2^{-1}$ holds (the condition that must be met if nutation is to be observed), we have $\Gamma \sim \chi(1+\chi/\sigma)$. The extrapolation of (9) into the region of very low χ yields the decay rate $(2T_2)^{-1}$ of irreversible relaxation, which can be found from electromagnetic echo-signal experiments. In the opposite case, i.e., when $\chi \geqslant \sigma(\langle V(t_m)\rangle) \to 1$, we have

$$\Gamma = \frac{\sigma}{\chi}\sqrt{\chi^2 + \frac{1}{4T_2^2}} + \frac{\chi^2 + 1/4T_2^2}{\chi} = \frac{\sigma}{\chi}\left[\int_0^\infty J_0(\chi t)\exp\left(-\frac{t}{2T_2}\right)dt\right]^{-1} + \left[\int_0^\infty \sin\chi t\,\exp\left(-\frac{t}{2T_2}\right)dt\right]^{-1}. \quad (10)$$

In other words, the roles played by the different components of the total integral decay rate change: the second term provides the main contribution to decay, while the contribution of the first term is determined by the value of the ratio σ/χ. If $\chi \geqslant (2T_2)^{-1}$, we have $\Gamma \simeq \chi(1+\sigma/\chi)$.

The above estimates can easily be verified by directly calculating the integral decay rate for signal (4) without allowing for irreversible relaxation:

$$\Gamma_+ = \langle V(t_m)\rangle\left[\chi\int_0^\infty t\,dt\int_0^1 \exp(-\sigma x t)J_0(\chi t\sqrt{1-x^2})dx\right]^{-1} = \langle V(t_m)\rangle(\sigma+\chi), \quad (11)$$

which can formally be interpreted as the result of calculating the damping constant of an exponential function $\exp[-(\sigma+\chi)t]$ normalized to the factor $\langle V(t_m)\rangle$. We can also find the rate of integral decay due to the coherent suppression mechanism:

$$\Gamma_1 = \left[\int_0^\infty J_0(\chi t)dt\right]^{-1} = \chi, \quad (12)$$

which can also be interpreted as the damping constant of an exponential function. In this representation, the expression (8) for the integral decay rate can be approximated by

$$\Gamma \simeq \frac{1}{2T_2} + \langle V(t_m)\rangle(\sigma+\chi). \quad (13)$$

This makes it possible to reproduce the procedure, done in Ref. 2, of distinguishing the oscillating part of the response described by a Bessel function by simply subtracting (12) from (13). If we subtract (11) from (13), we arrive at the natural result of Bloch's theory: in the absence of inhomogeneous broadening, the decay of the nutations is due only to irreversible relaxation.

Thus, allowance for the finiteness of the inhomogeneous linewidth enhances the decay of nutations. The above estimates show that for small values of χ the decay is a quadratic function of the amplitude, while for large values of χ it asymptotically tends to σ.

We now return to the experiment described in Ref. 2, where nutation decay was studied in connection with the second harmonic. For this case the above expressions must be modified, which is trivial, as shown in Ref. 7: χ must be replaced by $\chi J_1(\chi/\omega) = \tilde{\chi}$, where ω is the carrier frequency of the driving radiation.

3. RESULTS AND DISCUSSION

In order to compare the results with the experimental data of Ref. 2 we calculated the integral decay rates $\Gamma_2 = \Gamma - \Gamma_1$ and $\Gamma - \Gamma_+$ as functions of the Rabi frequency. Figure 3 depicts the data taken from Ref. 2 and our results on nutation decay in samples No. 1 ($T_2 = 120\times10^{-6}$ s and $\sigma = 0.125\times10^6\times2\pi$ Hz) and No. 2 ($T_2 = 8.7\times10^{-6}$ s and $\sigma = 0.625\times10^6\times2\pi$ Hz) studied in Ref. 2. We see that when the field amplitude is moderate, the theoretical results are in satisfactory agreement with the experimental data, and the region where this is true is wider for sample No. 2 than for sample No. 1. The explanation is that the inhomogeneous linewidth of sample No. 2 is five times larger than of sample No. 1. Hence for the second sample the broad-line approximation $\tilde{\chi}/\sigma \leqslant 1$ holds inn a wider interval of field amplitudes than for the first sample. In the case $\tilde{\chi}/\sigma \leqslant 1$ the theoretical integral decay rate exceeds the experimental rate, and the excess can be estimated if we employ the fact that both theoretical curves $\Gamma = \Gamma(\tilde{\chi})$ have linear sections. The slopes of these linear sections are $K_1 = 5.5\times10^{-2}$ and $K_2 = 6.45$

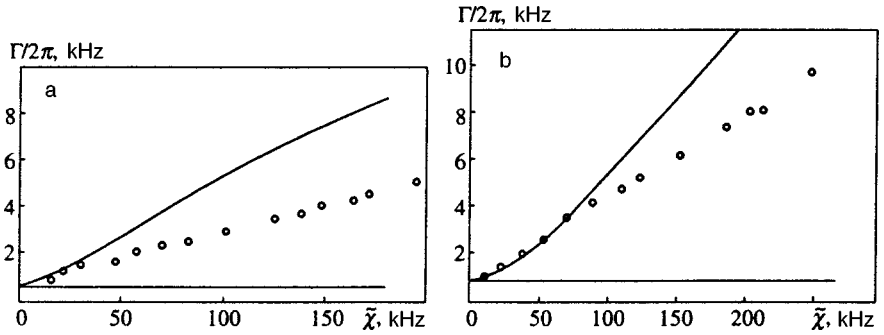


FIG. 3. Field dependence of the integral nutation-signal decay rate in samples No. 1 (a) and No. 2 (b). The curves represent the theoretical results and the open circles, the experimental data of Boscaino *et al.*[2]

$\times 10^{-2}$, respectively. These values are approximately twice as large as the experimental values[2] $K_1=(2.4\pm 0.1)\times 10^{-2}$ and $K_2=(3.3\pm 0.2)\times 10^{-2}$, respectively.

Thus, in contrast to the statement made by Boscaino *et al.*,[2] the additional (in comparison to the decay of the zeroth-order Bessel function) χ-dependent nutation-signal decay they detected can be described qualitatively by Bloch equations and is due to the finite value of the inhomogeneous linewidth. Boscaino *et al.*[2] did not take this fact into account, since they used the broad-line approximation in the entire range of values of $\tilde{\chi}$. In the light of our results, it must be noted that in estimating such a subtle effect as nutation decay the broad-line approximation should be used very cautiously. Our results were obtained by calculating the integral decay rate, while in Ref. 2 the decay was found over a finite interval of the evolution of the response. Hence theory and experiment agree here on the qualitative level. In a recent paper, Shakhmuratov *et al.*[8] attempted to explain the ''anomalous'' nutation decay detected by Boscaino *et al.*[2] using the stochastic model of the driving field. However, since in Ref. 2 the contribution to decay of the finite value of the inhomogeneous linewidth was not separated, we believe such an attempt was premature. To establish the full role that inhomogeneous line broadening plays in the response's decay it is advisable to carry out experiments in nutation decay in the one-photon regime. In such experiments the range of possible values of χ can be widened significantly and nutations can be studied for $\chi/\sigma>1$. Moreover, doing such an experiment will make it possible to exclude various distortions inherent in the two-photon method. Obviously, the results of the present paper hold not only for a magnetic resonance but also for an optical resonance.

4. CONCLUSION

We have established that within the scope of Bloch equations the inhomogeneous broadening of the spectral line not only coherently suppresses the nutation signal in a two-level quantum system but also causes exponential response decay, which depends on the ratio of the Rabi frequency to the inhomogeneous linewidth. At moderate field amplitudes the decay constant depends on the Rabi frequency according to a quadratic law, but in strong fields it asymptotically tends to a value equal to the inhomogeneous linewidth.

Thus, within the present approach we have given an explanation of the ''anomalous'' EPR-nutation decay detected in weak fields in quartz by Boscaino *et al.*[2]

This work was partially supported by the Belorussian Fund for Basic Research.

*)E-mail: lttt@ifttp.bas-net.by

[1] V. M. Akulin and N. V. Karlov, *Intense Resonant Interactions in Quantum Electrodynamics*, Springer, Berlin (1991).
[2] R. Boscaino, F. M. Gelardi, and J. P. Corb, Phys. Rev. B **48**, 7077 (1993).
[3] G. G. Fedoruk, Zh. Éksp. Teor. Fiz. **111**, 1207 (1997) [JETP **84**, 665 (1997)].
[4] H. C. Torrey, Phys. Rev. **76**, 1059 (1949).
[5] V. S. Kuz'min, Fiz. Tverd. Tela (St. Petersburg) **38**, 2704 (1996) [Phys. Solid State **38**, 1483 (1996)].
[6] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Sums, Series and Products*, Academic Press, New York (1980).
[7] V. S. Kuz'min and A. N. Yashin, Opt. Spektrosk. **62**, 1312 (1987) [Opt. Spectrosc. **62**, 775 (1987)].
[8] R. N. Shakhmuratov, F. M. Gelardi, and M. Cannas, Phys. Rev. Lett. **79**, 2963 (1997).

Translated by Eugene Yankovsky

Photon cooperative effect in resonance spectroscopy

B. A. Veklenko*)

Moscow Energy Institute, 111250 Moscow, Russia
(Submitted 10 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 492–510 (August 1998)

A systematic method is proposed for calculating the density matrix of subsystems interacting with their environment under conditions of thermodynamic equilibrium. The density matrix of photons resonantly interacting with a surrounding gas is calculated. It is shown that use of the Gibbs distribution allows one to completely eliminate inelastic processes from the calculations. A correct account of photon–photon correlators indicates the presence of new cooperative effects. A new branch of the polariton spectrum is predicted, which is due to the presence of excited atoms in the medium. With the help of the density matrix the mean filling numbers of the photon modes are calculated. In terms of wavelengths, we have obtained a generalization of the Planck formula which accounts for photon cooperative phenomena. The manifestation of these effects in kinetic processes is discussed. © *1998 American Institute of Physics.* [S1063-7761(98)01008-7]

1. INTRODUCTION

The kinetic phenomena in a medium consisting of atoms interacting with an electromagnetic field are quite varied. Most of them can be described with the help of second-order two-time Green's functions.[1–3] But one must also deal with so-called cooperative effects, which are not encompassed by this theory. A striking example is the Dicke effect (superradiance),[4] involving mutually-induced spontaneous emission from a few atoms. To describe this effect, it is necessary at a minimum to take account of fourth-order correlators $\langle \hat{\psi}^+ \hat{\psi}^+ \hat{\psi}\hat{\psi}\rangle$, where $\hat{\psi}$ is the field operator of the atom field.[5] An extensive literature is dedicated to the Dicke effect.[6–8] Interest in it has grown[9–11] in connection with the study of squeezed states of the electromagnetic field. For a large number of atoms the Dicke effect admits, as a rule, a quasiclassical description.[7]

We direct our attention to the existence of cooperative effects of a different kind. These cooperative effects are associated with correlators $\langle \hat{\alpha}^+_{\mathbf{k}\lambda} \hat{\alpha}^+_{\mathbf{k}\lambda} \hat{\alpha}_{\mathbf{k}\lambda} \hat{\alpha}_{\mathbf{k}\lambda}\rangle$, where $\hat{\alpha}_{\mathbf{k}\lambda}$ ($\hat{\alpha}^+_{\mathbf{k}\lambda}$) is the annihilation (creation) operator of a photon in a mode characterized by wave vector $\mathbf{k}$ and polarization index λ. Such effects are due to correlations between photons in real or virtual states. One manifestation of the photon cooperative effect is the specific character of the reflection of resonant radiation from excited media.[12,13] But this phenomenon does not admit a quasiclassical description. For light that has passed through a thin layer of atoms, the effect disappears.[14] Such an asymmetry forces us to carry out a more complete description. We correspondingly restrict the discussion to the state of thermodynamic equilibrium, whose theory is relatively simple. Against this background the behavior of interest to us stands out more clearly. The investigated cooperative effect is manifested here, in particular, in the form of an additional branch of the energy spectrum, whose presence allows us to suppose the existence of new effects in the kinetic phenomena.

Our approach was based on a modification of the method of Γ operators,[13] which automatically take into account all possible correlators of the electromagnetic field while preserving all of the attractive properties of the Green's function method, the Dyson equation, spectral representations, etc. The proposed formalism automatically distinguishes coherent scattering processes, in which the state of the medium does not change, from incoherent scattering processes, in which it does. It has been rigorously shown that the use of the Gibbs distribution to describe the interacting atom + field system allows one to omit all the incoherent processes. This circumstance, which does not hold in kinetic theory, simplifies things extraordinarily.

This work also pursues a different goal. It demonstrates that the proposed formalism allows one to effectively generalize the Gibbs distribution for subsystems interacting with the surrounding medium. It can be imagined that this fact may be useful in the study of phase transitions and nondestructive quantum measurements.[15,16] Specifically, in this work we explicitly calculate the density matrix of the photons in the medium. For a vanishingly small interaction of the photons with the atoms it, of course, coincides with the Gibbs distribution. With the help of the density matrix we calculate the distribution of the mean photon number over modes. This distribution is a generalization of the Planck formula. Features arising from cooperative phenomena are noted.

2. THE MODEL

We consider a nonrelativistic gas consisting of atoms with one valence electron. Spin effects are neglected. We assign the field operator $\hat{\psi}(\mathbf{r},\mathbf{R})$ in the Schrödinger representation to a particular atom of the gas. Here $\mathbf{R}$ is the position vector of the center of gravity of a particular atom and $\mathbf{r}$ is the position vector of the valence electron. We assign the operator $\hat{A}^{\nu}(\mathbf{r})$ to the transverse electromagnetic field. As-

suming the interaction to be quasiresonant, i.e., $|k-\omega_{m\mu}| \ll k+\omega_{m\mu}$ ($\omega_{m\mu}$ is the frequency of the optical transition in the atoms), we adopt the following Schrödinger equation to describe the system ($\hbar=c=1$):

$$i\frac{\partial\Psi}{\partial t}=\hat{H}\Psi. \tag{1}$$

Here

$$\hat{H}=\hat{H}^0+\hat{H}', \quad \hat{H}^0=\hat{H}_a+\hat{H}_{\rm ph}, \quad \hat{H}_{\rm ph}=\sum_{\mathbf{k}\lambda}\hat{H}_{\mathbf{k}\lambda},$$

$$\hat{H}_{\mathbf{k}\lambda}=k\left(\hat{\alpha}^+_{\mathbf{k}\lambda}\hat{\alpha}_{\mathbf{k}\lambda}+\frac{1}{2}\right), \quad \hat{H}_a=\sum_{i\mathbf{p}}\varepsilon_i(p)\hat{b}^+_{i\mathbf{p}}\hat{b}_{i\mathbf{p}},$$

$$\hat{H}'=-\frac{e}{m}\int\hat{\psi}^+(\mathbf{r},\mathbf{R})\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}(\mathbf{r})\hat{\psi}(\mathbf{r},\mathbf{R})d\mathbf{r}\,d\mathbf{R},$$

$$\hat{\psi}(\mathbf{r},\mathbf{R})=\sum_{i\mathbf{p}}\psi_i(\mathbf{r}-\mathbf{R})e^{i\mathbf{p}\cdot\mathbf{R}}\frac{\hat{b}_{i\mathbf{p}}}{\sqrt{V}}, \quad \varepsilon_i(p)=\varepsilon_i+\frac{p^2}{2M},$$

ψ_i is the wave vector describing the internal structure of the atoms of the gas with energy ε_i, $\mathbf{p}$ is the momentum of the atom, M is the mass of the atomic radical (atom minus electron), $V=L_xL_yL_z$ is the normalization volume, $\hat{b}^+_{i\mathbf{p}}$ ($\hat{b}_{i\mathbf{p}}$) are the creation (annihilation) operators of the atom in the state $(i,\mathbf{p})$;

$$\hat{A}^\nu(\mathbf{r})=\sum_{\mathbf{k}\lambda}\hat{A}^\nu_{\mathbf{k}\lambda}(\mathbf{r}), \quad \hat{A}^\nu_{\mathbf{k}\lambda}(\mathbf{r})=\frac{e^\lambda_{\mathbf{k}\nu}}{\sqrt{2kV}}\hat{\alpha}_{\mathbf{k}\lambda}e^{i\mathbf{k}\cdot\mathbf{r}}+\text{H.c.},$$

$\mathbf{e}^\lambda_{\mathbf{k}}$ is the linear polarization vector of the photons ($\lambda=1,2$). In the absence of temperature degeneracy of the gas the statistical properties of the operators $\hat{b}^+_{i\mathbf{p}}$ and $\hat{b}_{i\mathbf{p}}$ are of little consequence. We assume that

$$[\hat{b}_{i\mathbf{p}},\hat{b}^+_{i'\mathbf{p}'}]=\delta_{ii'}\delta_{\mathbf{pp}'}.$$

3. QUALITATIVE ANALYSIS

The role of the photon–photon correlators is made clear when induced processes are correctly accounted for in the propagation of a quantized field in the medium. Equation (1) can be rewritten in integral form

$$\Psi=\Psi^0+i\frac{e}{m}\int_{-\infty}^{t}\exp[-i\hat{H}^0(t-t')]\hat{\psi}^+\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}\hat{\psi}\,d\mathbf{r}\,d\mathbf{R}\,dt',$$

where Ψ^0 describes the state of the system before the interaction of the field with the medium is switched on. Assuming the medium to be transparent, we are interested only in processes of elastic light scattering. We seek the solution of the equation in the form of an iterative series. We omit the terms proportional to the odd powers of the coupling constant, which are responsible for the nontransparency of the medium. Summing the remaining subseries is equivalent to solving the following integral equation:

$$\Psi=\Psi^0-\left(\frac{e}{m}\right)^2\int_{-\infty}^{t}dt'\int\exp[-i\hat{H}^0(t-t')]$$
$$\times\hat{\psi}^+\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}\hat{\psi}d\mathbf{r}'d\mathbf{R}'\int_{-\infty}^{t'}\exp[-i\hat{H}^0(t'-t'')]$$
$$\times\hat{\psi}^+\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}\hat{\psi}d\mathbf{r}''d\mathbf{R}''\Psi(t'')dt''.$$

Since the atoms of the medium remain in their initial state as a result of elastic scattering, only the following products of operators $\hat{\psi}$ of the original set remain in the latter equation in the dipole approximation as a consequence of the selection rules:

$$\hat{\psi}^+\hat{\psi}\hat{\psi}^+\hat{\psi}\propto\hat{b}^+_{i\mathbf{p}}\hat{b}_{i'\mathbf{p}'}\hat{b}^+_{i'\mathbf{p}'}\hat{b}_{i\mathbf{p}}.$$

Under the action of such an operator, the wave function of the system remains factored

$$\Psi=\chi\varphi,$$

if it was factored before the interaction of the electromagnetic field with the medium was switched on,

$$\Psi^0=\chi^0\varphi, \quad \varphi=\prod_{i\mathbf{p}}\hat{b}^+_{i\mathbf{p}}|0\rangle\exp[-i\varepsilon_i(p)t].$$

Only the operators of the states occupied by the atoms are switched on in the product. Here

$$\chi=\chi^0+\int\Delta^0_r(t-t')\hat{\mathscr{P}}_r(t'-t'')\chi(t'')dt'\,dt'', \tag{2}$$

where

$$\Delta^0_r(t)=-ie^{-i\hat{H}^0t}\theta(t),$$

$$\hat{\mathscr{P}}_r(t)=\int e^{-iEt}\hat{\mathscr{P}}_r(E)\frac{dE}{2\pi},$$

$$\hat{\mathscr{P}}_r(E)=\sum_{\mathbf{k}\lambda\mathbf{k}'\lambda'}[\hat{\alpha}^+_{\mathbf{k}'\lambda'}c^{\mathbf{k}\lambda}_r(E-\hat{H}_{\rm ph})$$
$$\times\hat{\alpha}_{\mathbf{k}\lambda}+\hat{\alpha}_{\mathbf{k}\lambda}a^{\mathbf{k}\lambda}_r(E-\hat{H}_{\rm ph})\hat{\alpha}^+_{\mathbf{k}\lambda}],$$

and $\theta(t)$ is the Heaviside step function. For the structural coefficients we have the following expressions:

$$c^{\mathbf{k}\lambda}_r(E)=\sum_{ii'\mathbf{p}}\left|\frac{P^\lambda_{i'i}(\mathbf{k})}{\sqrt{2kV}}\right|^2\frac{N_i(\mathbf{p})\delta(\mathbf{p}-\mathbf{p}'+\mathbf{k})}{E+\varepsilon_i(p)-\varepsilon_{i'}(p')+i0},$$

$$a^{\mathbf{k}\lambda}_r(E)=\sum_{ii'\mathbf{p}}\left|\frac{P^\lambda_{ii'}(\mathbf{k})}{\sqrt{2kV}}\right|^2\frac{N_i(\mathbf{p})\delta(\mathbf{p}-\mathbf{p}'-\mathbf{k})}{E+\varepsilon_i(p)-\varepsilon_{i'}(p')+i0}.$$

$N_i(\mathbf{p})$ denotes the filling number of the ith atomic state. In the dipole approximation

$$P^\lambda_{ii'}(\mathbf{k})=\frac{e}{m}\int\psi^*_i(\boldsymbol{\rho})\mathbf{e}^\lambda_{\mathbf{k}}\cdot\hat{\mathbf{p}}\psi_{i'}(\boldsymbol{\rho})d\boldsymbol{\rho}.$$

For simplicity we neglect the Doppler effect and restrict the treatment to the two-level approximation for the atoms, denoting the Zeeman sublevels of their excited state by the subscript m, and their unexcited state by the subscript μ. Now

$$\hat{\mathscr{P}}_r(E)|N_{\mathbf{k}\lambda}\rangle=\sum_{m\mu}\left|\frac{P^{\lambda}_{m\mu}(\mathbf{k})}{\sqrt{2kV}}\right|^2 \times\left[\frac{N_\mu N_{\mathbf{k}\lambda}}{E-k(N_{\mathbf{k}\lambda}-1/2)-\omega_{m\mu}+i0}+\frac{N_m(N_{\mathbf{k}\lambda}+1)}{E-k(N_{\mathbf{k}\lambda}+3/2)+\omega_{m\mu}+i0}\right]|N_{\mathbf{k}\lambda}\rangle. \tag{3}$$

The operator (3), defining the evolution of the field, possesses the following properties. First of all, it is not algebraically related to the dielectric constant, which for excited rarefied media depends[1,2] on $N_\mu - N_m$. The dispersion relation corresponding to Eq. (2) has three branches, one of which disappears if $N_m=0$. This branch arises as a result of virtual induced emission processes. The last term of expression (3)—the term responsible for this branch—describes inverse scattering processes, which are well known in quantum electrodynamics. For processes of this kind, according to Eq. (3) the excited atom at first is induced to emit an additional photon into the $(\mathbf{k},\lambda)$ mode and only later does it absorb one of the photons of the field inducing this process. In the virtual state the effect per photon is greater than in real states before and after scattering. All the photons are correlated. Thus, the cooperation of the photons is closely associated with processes of induced emission of the atoms. This fact is unknown to semiclassical theory. A correct description of virtual induced emission is nowhere to be found in the technique of closed equations for the mean quantities.[1–3]

Note that in Eq. (2) we have omitted incoherent processes, which alter the state of the atoms as a result of scattering. Such processes introduce their own contribution to kinetic phenomena. Below we will show that based on the Gibbs distribution under conditions of thermodynamic equilibrium, incoherent processes can be completely excluded from the calculations. The equilibrium distribution function in this case is completely determined by the operator (3).

4. METHOD OF SPLIT Γ OPERATORS

To account for photon correlators in an arbitrary mode $(\mathbf{k},\lambda)$ we proceed as follows. Let the filling of the mode be characterized by the number $N_{\mathbf{k}\lambda}$. We introduce the auxiliary Γ space with the generating vector $\rangle^0_\Gamma$, in which the creation operators $\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})$ are defined so that the wave function of the state consisting of $N_{\mathbf{k}\lambda}$ photons in this space has the form

$$\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})\rangle^0_\Gamma.$$

We define the annihilation operator $\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})$ in a corresponding way. Since powers of these operators exceeding unity do not arise in the theory, their specific commutation relations have no effect on the final results. We assume that

$$[\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda}),\hat{\mathfrak{A}}^+(N'_{\mathbf{k}'\lambda'})]=\delta(N_{\mathbf{k}\lambda},N'_{\mathbf{k}\lambda})\,\delta_{\mathbf{k}\mathbf{k}'}\,\delta_{\lambda\lambda'}.$$

We may point out one possible realization of the Γ space. We apply the function

$$\mathscr{H}_n(\zeta(N_{\mathbf{k}\lambda}))\exp\left[-\frac{\xi^2(N_{\mathbf{k}\lambda})}{2}\right],$$

to the set $N_{\mathbf{k}\lambda}$, where $\mathscr{H}_n$ are the Hermite polynomials. Now

$$\rangle^0_\Gamma=\prod_{\mathbf{k}\lambda}\prod_{N_{\mathbf{k}\lambda}}\mathscr{H}_0(\xi(N_{\mathbf{k}\lambda}))\exp\left[-\frac{\xi^2(N_{\mathbf{k}\lambda})}{2}\right],$$

$$\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})=\frac{1}{\sqrt{2}}\left(\xi(N_{\mathbf{k}\lambda})+\frac{\partial}{\partial\xi(N_{\mathbf{k}\lambda})}\right),$$

$$\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})=\frac{1}{\sqrt{2}}\left(\xi(N_{\mathbf{k}\lambda})-\frac{\partial}{\partial\xi(N_{\mathbf{k}\lambda})}\right).$$

The wave function $\hat{\mathfrak{A}}^+(N'_{\mathbf{k}\lambda})\rangle^0_\Gamma$ is the product $\rangle^0_\Gamma$ given above, in which the subscript n on the function with argument $N'_{\mathbf{k}\lambda}$ is equal to unity. Since the set $N'_{\mathbf{k}\lambda}$ cannot be encountered in the theory in more than one realization, none of the subscripts $\ldots,n,\ldots$ can exceed unity. The merit of the formalism lies in precisely this property. Now it is clear that any power of the annihilation operator greater than one [i.e., $\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda}))^n$ for $n>1$] annihilates any physical state and extraordinarily simplifies the formalism. No products of the operators $\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})$ arise in the theory. In other words, complicated correlators do not arise and the need to cut them off in the calculations falls away. This fact allows the theory to take all the photon–photon correlators exactly into account.

Other realizations of the Γ representation are possible, leading to the same final results. Reference 13 used more general Γ operators $\hat{\mathfrak{A}}^+(\mathbf{N})$ corresponding to the total aggregate $\mathbf{N}=\ldots,N_{\mathbf{k}\lambda},\ldots$ of photons. In terms of the notation of the present paper

$$\hat{\mathfrak{A}}^+(\mathbf{N})=\prod_{\mathbf{k}\lambda}\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda}).$$

The use of simpler, one may say split, Γ operators $\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})$ makes it possible to represent the result of calculations in factored form in the manner of the Hartree representation in traditional quantum mechanics. If $\varphi(N_{\mathbf{k}\lambda}|\zeta_{\mathbf{k}\lambda})$ are the wave functions of a quantum oscillator, then the wave function of a free electromagnetic field can be represented in the form

$$\prod_{\mathbf{k}\lambda}\varphi(N_{\mathbf{k}\lambda}|\zeta_{\mathbf{k}\lambda}).$$

The unitary transformation from the standard representation of the filling numbers to the Γ representation is realized by the operator[13]

$$\hat{O}=\prod_{\mathbf{k}\lambda}\hat{\Phi}^+(\zeta_{\mathbf{k}\lambda})\rangle^0_\Gamma,\quad \hat{\Phi}(\zeta_{\mathbf{k}\lambda})=\sum_{N_{\mathbf{k}\lambda}}\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})\varphi(N_{\mathbf{k}\lambda}|\zeta_{\mathbf{k}\lambda}).$$

The wave function of an arbitrary free electromagnetic field in the Γ representation has the form

$$\prod_{\mathbf{k}\lambda}\sum_{N_{\mathbf{k}\lambda}}C(N_{\mathbf{k}\lambda})\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})\rangle^0_\Gamma, \tag{4}$$

here $C(N_{\mathbf{k}\lambda})$ are the expansion coefficients. Now it is obvious that

$$\int \hat{\Phi}^+(\zeta_{\mathbf{k}\lambda})\hat{\Phi}(\zeta_{\mathbf{k}\lambda})d\zeta_{\mathbf{k}\lambda}=\sum_{N_{\mathbf{k}\lambda}} \hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})=1. \tag{5}$$

In the Γ representation the Schrödinger equation has the form

$$i\frac{\partial\Psi_\Gamma}{\partial t}=\Bigg[\hat{H}_a+\sum_{\mathbf{k}\lambda}\int\hat{\Phi}^+(\zeta_{\mathbf{k}\lambda})\hat{H}_{\mathbf{k}\lambda}\hat{\Phi}(\zeta_{\mathbf{k}\lambda})d\zeta_{\mathbf{k}\lambda}-\frac{e}{m}\times\sum_{\mathbf{k}\lambda}\int\hat{\Phi}^+(\zeta_{\mathbf{k}\lambda})\hat{\psi}^+\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}_{\mathbf{k}\lambda}(\mathbf{r})\hat{\psi}\hat{\Phi}\times(\zeta_{\mathbf{k}\lambda})d\mathbf{r}\,d\mathbf{R}\,d\zeta_{\mathbf{k}\lambda}\Bigg]\Psi_\Gamma.$$

The mean value of any field operator $\hat{K}_{\mathbf{k}\lambda}$ is given by

$$\langle K_{\mathbf{k}\lambda}\rangle=\mathrm{Tr}\,\hat{K}_{\mathbf{k}\lambda}\rho_{\mathbf{k}\lambda},$$

where

$$\rho_{\mathbf{k}\lambda}=\rho(\zeta_{\mathbf{k}\lambda},\zeta'_{\mathbf{k}\lambda})=\langle\hat{\Phi}^+(\zeta'_{\mathbf{k}\lambda})\hat{\Phi}(\zeta_{\mathbf{k}\lambda})\rangle_\Gamma.$$

Here

$$\rangle_\Gamma=\Psi_\Gamma.$$

For systems in thermodynamic equilibrium, averaging can be performed simultaneously over quantum states and over the ensemble of systems. Therefore

$$\rho(\zeta_{\mathbf{k}\lambda},\zeta'_{\mathbf{k}\lambda})=\mathrm{Tr}\langle\hat{\Phi}^+(\zeta'_{\mathbf{k}\lambda})\hat{\Phi}(\zeta_{\mathbf{k}\lambda})\rho_H\rangle_\Gamma,$$

$$\rho_H=\exp\frac{\Omega-\hat{H}-\mu\hat{N}_a}{T},$$

where $\hat{N}_a$ is the number operator of the atoms, μ is their chemical potential, and Ω is the thermodynamic potential of the system.

If we transform to the arguments $N_{\mathbf{k}\lambda}$ with the help of the quantum oscillator functions, then making use of the diagonality of ρ under equilibrium conditions, we have

$$\rho(N_{\mathbf{k}\lambda},N_{\mathbf{k}\lambda})=\rho(N_{\mathbf{k}\lambda})=\mathrm{Tr}\langle\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})\rho_H\rangle_\Gamma. \tag{6}$$

and it follows from Eqs. (5) and (6) that

$$\sum_{N_{\mathbf{k}\lambda}}\rho(N_{\mathbf{k}\lambda})=1. \tag{7}$$

5. METHOD OF TEMPERATURE GREEN'S FUNCTIONS

It is convenient to calculate the desired construction (6) by the method of Matsubaro Green's functions.[1] To this end we introduce the function

$$\mathscr{D}(N_{\mathbf{k}\lambda},\tau,N'_{\mathbf{k}\lambda},\tau')=-\mathrm{Tr}\,\rho_H\hat{T}_\tau\check{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau)\check{\tilde{\mathfrak{A}}}(N'_{\mathbf{k}\lambda},\tau'), \tag{8}$$

where $\hat{T}_\tau$ is the time-ordering operator in the parameter τ. The trace symbol (Tr) here represents averaging in both the quantum and statistical senses. The Heisenberg Γ operators are constructed as follows:

$$\check{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau)=\exp[(\hat{H}-\mu\hat{N}_a)\tau]\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})\exp[-(\hat{H}-\mu\hat{N}_a)\tau],$$

$$\check{\tilde{\mathfrak{A}}}(N_{\mathbf{k}\lambda},\tau)=\exp[(\hat{H}-\mu\hat{N}_a)\tau]\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})\times\exp[-(\hat{H}-\mu\hat{N}_a)\tau], \tag{9}$$

the parameter τ is assumed to be positive. The desired density matrix ρ is found by taking the limit

$$\mathscr{D}\underset{\tau'\to\tau+0}{\to}\delta(N_{\mathbf{k}\lambda},N'_{\mathbf{k}\lambda})\rho(N_{\mathbf{k}\lambda}). \tag{10}$$

6. THE INTERACTION REPRESENTATION

To calculate the temperature Green's function (8), we transform to the Matsubaro interaction representation. We define in the interaction representation the operators $\check{\tilde{\mathfrak{A}}}(N_{\mathbf{k}\lambda},\tau)$ and $\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda,\tau})$ with the help of formulas (9), in which the operator $\hat{H}$ is replaced by the operator $\hat{H}^0$. Re-expressing expression (8) in terms of these newly introduced operators is effected in the standard way[1] and leads to the result

$$\mathscr{D}(N_{\mathbf{k}\lambda},\tau,N'_{\mathbf{k}\lambda},\tau')=-Q^{-1}\times\left\langle\hat{T}_\tau\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau)\check{\tilde{\mathfrak{A}}}(N'_{\mathbf{k}\lambda},\tau')\hat{\mathfrak{S}}\left(\frac{1}{T}\right)\right\rangle^0. \tag{11}$$

The angular brackets here denote averaging both in the quantum sense over states of the noninteracting atomic and electromagnetic fields and in the standard sense with the weight

$$\rho^0=\exp\left(\frac{\Omega^0-\hat{H}^0-\mu\hat{N}_a}{T}\right),$$

Ω^0 is the thermodynamic potential in the absence of interaction

$$Q=\exp\left[-\frac{\Omega-\Omega^0}{T}\right]=\left\langle\hat{\mathfrak{S}}\left(\frac{1}{T}\right)\right\rangle^0.$$

As for the operator $\hat{\mathfrak{S}}$, it is equal to

$$\hat{\mathfrak{S}}(\tau)=\hat{T}_\tau\exp\left[\int_0^\tau\hat{H}'(\tau')d\tau'\right],$$

where $\hat{H}'(\tau)$ is the interaction operator in the interaction representation. It follows from the definition that

$$\frac{\partial}{\partial\tau}\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau)=[\hat{H}^0-\mu\hat{N}_a,\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau)]=-k(N_{\mathbf{k}\lambda}+1/2)\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau).$$

Therefore

$$\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau)=\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})\exp[-k(N_{\mathbf{k}\lambda}+1/2)\tau].$$

Analogously

$$\check{\tilde{\mathfrak{A}}}(N_{\mathbf{k}\lambda},\tau)=\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})\exp[k(N_{\mathbf{k}\lambda}+1/2)\tau].$$

7. FREE-FIELD GREEN'S FUNCTIONS

First, let us consider the difference between the time-ordered and normal operator products:

$$(\hat{T}_\tau-\hat{N})\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau)\hat{\bar{\mathfrak{A}}}(N'_{\mathbf{k}\lambda},\tau')$$
$$=-\delta(N_{\mathbf{k}\lambda},N'_{\mathbf{k}\lambda})\Delta^0(N_{\mathbf{k}\lambda},\tau-\tau'),$$

where

$$\Delta^0(N_{\mathbf{k}\lambda},\tau-\tau')=-[\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau)\hat{\bar{\mathfrak{A}}}(N'_{\mathbf{k}\lambda},\tau')]\theta(\tau-\tau')$$
$$=-\theta(\tau-\tau')\exp[-k(N_{\mathbf{k}\lambda}+1/2)(\tau-\tau')]. \tag{12}$$

If there is no interaction in the system, then $\hat{\mathfrak{S}}(1/T)=1$ and

$$\mathscr{D}^0=-\{[1+\rho^0(N_{\mathbf{k}\lambda})]\theta(\tau-\tau')+\rho^0(N_{\mathbf{k}\lambda})\theta(\tau'-\tau)\}$$
$$\times\delta(N_{\mathbf{k}\lambda},N'_{\mathbf{k}\lambda})\exp[-k(N_{\mathbf{k}\lambda}+1/2)(\tau-\tau')].$$

Here

$$\rho^0(N_{\mathbf{k}\lambda})=\langle\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})\rangle^0$$
$$=\exp\left[\frac{\Omega^0_{\mathbf{k}\lambda}-k(N_{\mathbf{k}\lambda}+1/2)}{T}\right],$$

where

$$\exp\left(\frac{\Omega^0_{\mathbf{k}\lambda}}{T}\right)=\frac{e^{k/2T}}{1-e^{-k/T}}.$$

Comparing the last three formulas allows us to write

$$\mathscr{D}^0=\Delta^0-\rho^0(N_{\mathbf{k}\lambda},\tau-\tau'),$$
$$\rho^0(N_{\mathbf{k}\lambda},\tau-\tau')=\rho^0(N_{\mathbf{k}\lambda})\exp[-k(N_{\mathbf{k}\lambda}+1/2)(\tau-\tau')]. \tag{13}$$

Since now

$$\rho^0(N_{\mathbf{k}\lambda},-1/T)=\exp(\Omega^0_{\mathbf{k}\lambda}/T),$$

it follows that

$$\rho^0(N_{\mathbf{k}\lambda},\tau-\tau')=-\Delta^0(N_{\mathbf{k}\lambda},\tau-\tau')\rho^0(N_{\mathbf{k}\lambda})$$
$$=-\Delta^0(N_{\mathbf{k}\lambda},\tau-\tau'+1/T)\rho^0(N_{\mathbf{k}\lambda},-1/T). \tag{14}$$

Analogously

$$\rho^0(N_{\mathbf{k}\lambda},\tau-\tau')=-\Delta^0(N_{\mathbf{k}\lambda},\tau)\rho^0(N_{\mathbf{k}\lambda},-\tau')$$
$$=-\rho^0(N_{\mathbf{k}\lambda},\tau-1/T)\Delta^0(N_{\mathbf{k}\lambda},1/T-\tau'). \tag{15}$$

For the Green's function of a free atomic field we have

$$G(X,X')=-\langle\hat{T}_\tau\hat{\psi}(X)\hat{\bar{\psi}}(X')\rangle^0,\quad X=\{\mathbf{r},\mathbf{R},\tau\}.$$

Using the interaction representation

$$\hat{\psi}(X)=\sum_{i\mathbf{p}}\psi_i(\mathbf{r}-\mathbf{R})\frac{\hat{b}_{i\mathbf{p}}}{\sqrt{V}}\exp[i\mathbf{p}\cdot\mathbf{R}-\varepsilon_i(p)\tau],$$

$$\hat{\bar{\psi}}(X)=\sum_{i\mathbf{p}}\psi_i^*(\mathbf{r}-\mathbf{R})\frac{\hat{b}^+_{i\mathbf{p}}}{\sqrt{V}}\exp[-i\mathbf{p}\cdot\mathbf{R}+\varepsilon_i(p)\tau],$$

in the absence of temperature degeneracy we find

$$G(X,X')=-\frac{1}{V}\sum_{j\mathbf{p}}\psi_j(\mathbf{r}-\mathbf{R})\psi_j^*(\mathbf{r}'-\mathbf{R}')\exp[i\mathbf{p}\cdot(\mathbf{R}-\mathbf{R}')]$$
$$\times\int_{-\infty}^{\infty}\delta_\gamma(E-\varepsilon_i(p))\exp[-E(\tau-\tau')]$$
$$\times dE[\theta(\tau-\tau')+N_j(\mathbf{p})\theta(\tau'-\tau)]. \tag{16}$$

Here $N_j(\mathbf{p})$ are the mean filling numbers of the state $(j,\mathbf{p})$, and δ_γ is the usual Dirac delta-function. But if we take account of the interaction of the atoms of the medium with foreign particles (the reservoir), which leads to a broadening of the energy spectrum, then the Dirac delta-function must be replaced by the Lorentzian

$$\delta_\gamma(E-\varepsilon)=-\frac{1}{2\pi i}\left(\frac{1}{E-\varepsilon+i\gamma/2}-\frac{1}{E-\varepsilon-i\gamma/2}\right),$$

where γ is the width of the energy level ε. The lower integration limit in Eq. (16) should be taken as large as is convenient, but finite.

8. SECOND-ORDER PERTURBATION THEORY

In the interaction representation

$$\hat{\Phi}(\zeta_{\mathbf{k}\lambda},\tau)=\sum_{N_{\mathbf{k}\lambda}}\varphi(N_{\mathbf{k}\lambda}|\zeta_{\mathbf{k}\lambda})\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})\exp(-kN_{\mathbf{k}\lambda}\tau),$$

$$\hat{\bar{\Phi}}(\zeta_{\mathbf{k}\lambda},\tau)=\sum_{N_{\mathbf{k}\lambda}}\varphi(N_{\mathbf{k}\lambda}|\zeta_{\mathbf{k}\lambda})\hat{\mathfrak{A}}^+(N_{\mathbf{k}\lambda})\exp(kN_{\mathbf{k}\lambda}\tau).$$

Therefore

$$\hat{H}'(\tau)=-\frac{e}{m}\sum_{\mathbf{k}\lambda N_{\mathbf{k}\lambda}N'_{\mathbf{k}\lambda}}\int\hat{\bar{\mathfrak{A}}}(N_{\mathbf{k}\lambda},\tau)\hat{\bar{\psi}}\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}_{\mathbf{k}\lambda}\hat{\psi}\hat{\mathfrak{A}}$$
$$\times(N'_{\mathbf{k}\lambda},\tau)d\mathbf{r}\,d\mathbf{R}.$$

The operator $\hat{\mathbf{A}}_{\mathbf{k}\lambda}$ here is understood to mean the expression

$$\int\varphi(N_{\mathbf{k}\lambda}|\zeta_{\mathbf{k}\lambda})\hat{\mathbf{A}}_{\mathbf{k}\lambda}(\mathbf{r})\varphi(N'_{\mathbf{k}\lambda}|\zeta_{\mathbf{k}\lambda})d\zeta_{\mathbf{k}\lambda}. \tag{17}$$

In essence, the transformation (17) effects only a unitary change in the arguments. For this reason we retain the old notation for the operator $\hat{\mathbf{A}}_{\mathbf{k}\lambda}$ in the new arguments.

Let us write out expression (11) in second-order perturbation theory. From the arguments $\zeta_{\mathbf{k}\lambda}$ we transform with the help of $\varphi(N_{\mathbf{k}\lambda}|\zeta_{\mathbf{k}\lambda})$ to the arguments $N_{\mathbf{k}\lambda}$. We have

$$Q\mathscr{D}=\mathscr{D}^0-\left\langle\hat{T}_\tau\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda},\tau)\hat{\bar{\mathfrak{A}}}(N'_{\mathbf{k}\lambda},\tau')\frac{1}{2!}\left(\frac{e}{m}\right)^2\right.$$
$$\left.\times\int\hat{\bar{\mathfrak{A}}}\hat{\bar{\psi}}\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}_{\mathbf{k}\lambda}\hat{\psi}\hat{\mathfrak{A}}dX_1\int\hat{\bar{\mathfrak{A}}}\hat{\bar{\psi}}\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}_{\mathbf{k}\lambda}\hat{\psi}\hat{\mathfrak{A}}dX_2\right\rangle,$$

$$dX=d\mathbf{r}\,d\mathbf{R}\,d\tau.$$

For brevity, we have dropped the obvious arguments from the operators $\hat{\mathfrak{A}}$ and $\hat{\psi}$. To simplify the product of $\hat{\psi}$ operators, we use a thermodynamic variant of Wick's theorem.[1] In other words, we assume that in the thermodynamic limit the

higher correlators are expressed in terms of the lower ones. Possible pairings of operators are indicated by a horizontal bracket:

$$\overbrace{\hat{\psi}\hat{\psi}}=-G(X,X').$$

To simplify the products of operators $\hat{\mathfrak{A}}$ we make use of an algebraic theorem due to Wick[17] and the fact that in the physical states (4) the identity $\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})\hat{\mathfrak{A}}(N_{\mathbf{k}\lambda})=0$ holds. Thus, out of all $\hat{N}$ possible products, only those ''survive'' that contain only one annihilation operator corresponding to the mode $(\mathbf{k},\lambda)$. For this reason, the function $\rho^0(N_{\mathbf{k}\lambda})$ can appear only to the first power in any term.

The use of Wick's algebraic theorem does not introduce any approximations or simplifications. For this reason, the photon–photon correlators are taken into account in this technique exactly. Now

$$\begin{aligned}Q\mathscr{D}=\mathscr{D}^0-\left(\frac{e}{m}\right)^2\sum_{\nu_1\nu_2}\int \hat{p}^{\nu_1}G(X,X')\hat{p}^{\nu_2}G(X',X)\\ \times[-\Delta^0\hat{A}^{\nu_1}_{\mathbf{k}\lambda}\Delta^0\hat{A}^{\nu_2}_{\mathbf{k}\lambda}\Delta^0+\rho^0\hat{A}^{\nu_1}_{\mathbf{k}\lambda}\Delta^0\hat{A}^{\nu_2}_{\mathbf{k}\lambda}\Delta^0\\ +\Delta^0\hat{A}^{\nu_1}_{\mathbf{k}\lambda}\rho^0\hat{A}^{\nu_2}_{\mathbf{k}\lambda}\Delta^0+\Delta^0\hat{A}^{\nu_1}_{\mathbf{k}\lambda}\Delta^0\hat{A}^{\nu_2}_{\mathbf{k}\lambda}\rho^0]dX'dX\end{aligned}$$

(the operators $\hat{p}^{\nu}$ act on the closest arguments of the G function). We have taken into account only those terms corresponding to connected Feynman diagrams. The last formula can be rewritten as follows:

$$Q\mathscr{D}=\Delta^0-\rho^0+\Delta^0\hat{\mathscr{P}}\Delta^0-\rho^0\hat{\mathscr{P}}\Delta^0-\Delta^0\hat{\mathscr{P}}\rho^0-\Delta^0\hat{\mathscr{P}}^{(n)}\Delta^0, \tag{18}$$

where

$$\begin{aligned}\hat{\mathscr{P}}=\left(\frac{e}{m}\right)^2\int \hat{\mathbf{p}}\cdot\hat{\mathbf{A}}_{\mathbf{k}\lambda}G(X,X')\Delta^0\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}_{\mathbf{k}\lambda}G(X',X)\\ \times d\mathbf{r}\,d\mathbf{r}'\,d\mathbf{R}\,d\mathbf{R}',\end{aligned} \tag{19}$$

$$\begin{aligned}\hat{\mathscr{P}}^{(n)}=\left(\frac{e}{m}\right)^2\int \hat{\mathbf{p}}\cdot\hat{\mathbf{A}}_{\mathbf{k}\lambda}G(X,X')\rho^0\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}_{\mathbf{k}\lambda}G(X',X)\\ \times d\mathbf{r}\,d\mathbf{r}'\,d\mathbf{R}\,d\mathbf{R}'.\end{aligned} \tag{20}$$

Expressions (19) and (20) are, respectively, the first terms of the expansions in the parameter $n\lambda\!\!\!^{-3}$, where $\lambda\!\!\!^{-}=2\pi/k$ and n is the concentration of the atoms. Therefore, the smallness of this parameter serves as a condition of the applicability of the above formulas.

9. GENERAL STRUCTURE OF THE PERTURBATION-THEORY SERIES

An account of the higher approximations in expression (18) leads to the result that Δ^0 is augmented by terms not containing $\rho^0(N_{\mathbf{k}\lambda})$ corresponding to the connected Feynman diagrams. Their sum with Δ^0 is denoted as Δ. This sum is multiplied by the sum of all possible loop diagrams, which is equal[1] to Q. Thus,

$$Q\mathscr{D}=Q\Delta-Q_{\mathbf{k}\lambda}\rho^{\mathrm{con}}.$$

The variable ρ^{con} denotes the sum of connected Feynman diagrams, each of which necessarily contains $\rho^0(N_{\mathbf{k}\lambda})$. This quantity, in turn, is multiplied by the sum of all possible loop diagrams $Q_{\mathbf{k}\lambda}$, in which, however, $\rho^0(N_{\mathbf{k}\lambda})$ is absent. Below we restrict the discussion to an account of linear processes of the interaction of light with matter. In other words, we assume that the operator $\hat{\mathscr{P}}$ does not depend on $\rho^0(N_{\mathbf{k}'\lambda'})$ for any $(\mathbf{k}',\lambda')$. As for the operator $\hat{\mathscr{P}}^{(n)}$, it depends linearly on $\rho^0(N_{\mathbf{k}\lambda})$. Under these conditions the contributions of all the Feynman diagrams are summed up in the following equations:

$$\Delta=\Delta^0+\Delta^0\hat{\mathscr{P}}\Delta, \tag{21}$$

$$\rho^{\mathrm{con}}=\rho^0+\rho^0\hat{\mathscr{P}}\Delta+\Delta^0\hat{\mathscr{P}}\rho^{\mathrm{con}}+\Delta^0\hat{\mathscr{P}}^{(n)}\Delta. \tag{22}$$

According to the limit (10) the desired matrix ρ is given by

$$\rho(N_{\mathbf{k}\lambda})=-\mathscr{D}=\frac{Q_{\mathbf{k}\lambda}}{Q}\rho^{\mathrm{con}},\quad \tau'\to\tau+0. \tag{23}$$

Equation (22) can be written in the form

$$(1-\Delta^0\hat{\mathscr{P}})\rho^{\mathrm{con}}=\rho^0(1+\hat{\mathscr{P}}\Delta)+\Delta^0\hat{\mathscr{P}}^{(n)}\Delta. \tag{24}$$

A direct check using Eq. (21) shows that

$$(1+\Delta\hat{\mathscr{P}})(1-\Delta^0\hat{\mathscr{P}})=1.$$

Now, after multiplication on the left by the operator $1+\Delta\hat{\mathscr{P}}$ Eq. (24) takes the form

$$\rho^{\mathrm{con}}=(1+\Delta\hat{\mathscr{P}})\rho^0(1+\hat{\mathscr{P}}\Delta)+\Delta\hat{\mathscr{P}}^{(n)}\Delta.$$

This equation serves as a basis for further discussion. It is convenient to write it in the form

$$\rho^{\mathrm{con}}=\rho^{(c)}+\rho^{(n)},$$

where

$$\rho^{(c)}=(1+\Delta\hat{\mathscr{P}})\rho^0(1+\hat{\mathscr{P}}\Delta), \tag{25}$$

$$\rho^{(n)}=\Delta\hat{\mathscr{P}}^{(n)}\Delta. \tag{26}$$

The matrix $\rho^{(c)}$ contains scattering processes, as a result of which the state of the scatterers does not vary (coherent channel). The matrix $\rho^{(n)}$ corresponds to processes that alter the state of the scattering system (incoherent channel). The function Δ^0 and along with it the operator $\hat{\mathscr{P}}$ have a ''retarding'' character. The operator Δ, according to Eq. (21), also possesses this property. Since the integration in expression (26) extends over the interval from zero to $1/T$,

$$\rho^{(n)}(\tau,1/T)=0,\quad \tau<1/T.$$

According to the definition (8), all the functions depend on τ and τ' in terms of differences. Consequently,

$$\rho^{\mathrm{con}}(\tau,1/T)=\rho^{(c)}(\tau-1/T),\quad \tau<1/T.$$

Making the substitution of arguments $\tau\to(1/T)+\tau-\tau'<1/T$, we have

$$\rho^{\mathrm{con}}(\tau-\tau')=\rho^{(c)}(\tau-\tau')\quad \text{for } \tau<\tau'$$

and

$$\rho^n(\tau-\tau')=0 \quad \text{for } \tau<\tau'. \tag{27}$$

In the lowest order of perturbation theory the polarization operator $\hat{\mathscr{P}}^{(n)}$ is given by formula (20). Substitution of this expression into Eq. (26) leads to a contradiction with Eq. (27). We arrive at the conclusion that the incoherent channel cannot be calculated in perturbation theory; therefore, we are required here to sum up infinite subsequences of Feynman diagrams. For our purposes, there is no need according to Eqs. (23) and (27) to calculate processes of the incoherent channel. Thus, the given variant, in essence, represents a manifestation of the fluctuation–dissipation theorem.

Let us consider the coherent channel. By virtue of the retarding nature of the propagator Δ, it follows from Eq. (25) that

$$\rho^{(c)}(0,\tau')=\rho^0(1+\hat{\mathscr{P}}\Delta),$$

$$\rho^{(c)}(\tau,1/T)=(1+\Delta\hat{\mathscr{P}})\rho^0 \quad \text{for } \tau<1/T.$$

Employing formulas (14) and (15) and Eq. (21) we can recast the above two relations in the following form:

$$\rho^{(c)}(0,\tau')=-\rho^0(0,1/T)\Delta(1/T,\tau'), \tag{28}$$

$$\rho^{(c)}(\tau,1/T)=-\Delta(\tau,0)\rho^0(0,1/T). \tag{29}$$

Taking the difference nature of the dependence of these functions on τ and τ' into account, Eqs. (27)–(29) lead to the general formula

$$\rho^{\mathrm{con}}(\tau-\tau')=-\Delta(\tau-\tau'+1/T)\rho^0(0,1/T), \quad \tau<\tau'. \tag{30}$$

According to Eqs. (23) and (30), the desired density matrix can be found from the expression

$$\rho=-\frac{1}{Z}\Delta\left(\frac{1}{T}\right), \quad Z^{-1}=\frac{Q_{\mathbf{k}\lambda}}{Q}\exp\left(\frac{\Omega^0_{\mathbf{k}\lambda}}{T}\right). \tag{31}$$

The quantity Z can be calculated by the graphical technique, but it is simpler to use the normalization condition (7).

10. THE PROPAGATOR Δ

According to Eqs. (31), it is necessary to calculate $\Delta(1/T)$. This function is found from Eq. (21), which for $\tau<1/T$, by virtue of the retarding properties of the functions entering into it, can be written in the form

$$\Delta=\Delta^0+\int_0^\tau\Delta^0(\tau-\tau')\int_0^{\tau'}\hat{\mathscr{P}}(\tau'-\tau'')\Delta(\tau'')d\tau'd\tau''.$$

For the Laplace transform we have the obvious result

$$\Delta(s)=\frac{\Delta^0(s)}{1-\Delta^0(s)\hat{\mathscr{P}}(s)}, \tag{32}$$

and according to Eq. (12)

$$\Delta^0(s)=-[s+k(N_{\mathbf{k}\lambda}+1/2)]^{-1}.$$

An explicit expression for $\hat{\mathscr{P}}(s)$ can be easily found from formulas (12), (16), and (19):

$$\hat{\mathscr{P}}(s)|N_{\mathbf{k}\lambda}\rangle=[(N_{\mathbf{k}\lambda}+1)a^{\mathbf{k}\lambda}(s)+N_{\mathbf{k}\lambda}c^{\mathbf{k}\lambda}(s)]|N_{\mathbf{k}\lambda}\rangle,$$

where the structural coefficients $c^{\mathbf{k}\lambda}$ and $a^{\mathbf{k}\lambda}$ are represented everywhere by analytical functions with boundary values on the real axis

$$c^{\mathbf{k}\lambda}(s\pm i0)=-\sum_{i_1i_2\mathbf{p}}\left|\frac{P^\lambda_{i_2i_1}(\mathbf{k})}{\sqrt{2kV}}\right|^2 \times\frac{N_{i_2}(\mathbf{p})}{s+\varepsilon_{i_1}(\mathbf{p}-\mathbf{k})-\varepsilon_{i_2}(p)+k(N_{\mathbf{k}\lambda}-1/2)\pm i\gamma/2},$$

$$a^{\mathbf{k}\lambda}(s\pm i0)=-\sum_{i_1i_2\mathbf{p}}\left|\frac{P^\lambda_{i_1i_2}(\mathbf{k})}{\sqrt{2kV}}\right|^2 \times\frac{N_{i_2}(\mathbf{p})}{s+\varepsilon_{i_1}(\mathbf{p}-\mathbf{k})-\varepsilon_{i_2}(p)+k(N_{\mathbf{k}\lambda}+3/2)\pm i\gamma/2},$$

where $\gamma=\gamma_{i_1}+\gamma_{i_2}$.

The resemblance of the above two formulas to expression (3) is obvious. The inverse Laplace transform

$$\Delta(\tau)=\int_{a-i\infty}^{a+i\infty}e^{\tau s}\Delta(s)\frac{ds}{2\pi i}$$

allows us to find the final result.

A spectral analysis, analogous to the spectral analysis of standard Green's functions,[1] shows that the propagator $\Delta(s)$ can be represented by an analytical function with singularities only on the real axis. Therefore

$$\Delta(\tau)=\int_{E_{\min}}^\infty e^{-\tau s}[\Delta(-s-i0)-\Delta(-s+i0)]\frac{ds}{2\pi i}. \tag{33}$$

The constant $E_{\min}$ coincides with the lower limit of the energy spectrum of the system as a whole. In practical calculations, the question of its magnitude is resolved individually in each case. As above, we use the two-level approximation for the atoms. We neglect the Doppler effect. In the dipole approximation we have

$$c^{\mathbf{k}\lambda}(-s\pm i0)=\frac{-c}{-s+\omega_{m\mu}+k(N_{k\lambda}-1/2)\pm i\gamma/2},$$

$$c=\sum_{m\mu}\frac{|P^\lambda_{m\mu}(\mathbf{k})|^2n_\mu}{2k},$$

$$a^{\mathbf{k}\lambda}(-s\pm i0)=\frac{-a}{-s-\omega_{m\mu}+k(N_{k\lambda}+3/2)\pm i\gamma/2},$$

$$a=\sum_{m\mu}\frac{|P^\lambda_{m\mu}(\mathbf{k})|^2n_m}{2k}.$$

To finish evaluating the coefficients $c^{\mathbf{k}\lambda}$ and $a^{\mathbf{k}\lambda}$, we can make use of the fact that

$$\sum_{m\mu}|P^\lambda_{m\mu}(\mathbf{k})|^2=\pi\gamma_r\frac{2j_m+1}{\omega_{m\mu}}.$$

Here j_m is the orbital quantum number and γ_r^{-1} is the radiative lifetime of the excited state of the atom. The zeros of the denominator of the integrand (32) form a spectrum with three branches. Therefore

$$\Delta(-s-i0)=\frac{[s-\omega_{m\mu}-k(N_{\mathbf{k}\lambda}-1/2)+i\gamma/2][s+\omega_{m\mu}-k(N_{\mathbf{k}\lambda}+3/2)+i\gamma/2]}{(s-s_0)(s-s_1)(s-s_2)}.$$

To first order in the expansions in the concentrations n_μ and n_m we have

$$s_0=s_0^{(0)}+\Delta s_0,\quad s_1=s_1^{(0)}+\Delta s_1,\quad s_2=s_2^{(0)}+\Delta s_2,$$

where

$$s_0^{(0)}=k\left(N_{\mathbf{k}\lambda}+\frac{1}{2}\right),\quad s_1^{(0)}=\omega_{m\mu}+k\left(N_{\mathbf{k}\lambda}-\frac{1}{2}\right)-i\frac{\gamma}{2},$$

$$s_2^{(0)}=-\omega_{m\mu}+k\left(N_{\mathbf{k}\lambda}+\frac{3}{2}\right)-i\frac{\gamma}{2},$$

$$\Delta s_0=N_{\mathbf{k}\lambda}(\eta_\mu-\eta_m)(k-\omega_{m\mu})-i\frac{\gamma}{2}N_{\mathbf{k}\lambda}(\eta_\mu+\eta_m)$$
$$-\eta_m\left(k-\omega_{m\mu}+i\frac{\gamma}{2}\right),$$

$$\Delta s_1=-N_{\mathbf{k}\lambda}\left(k-\omega_{m\mu}-i\frac{\gamma}{2}\right)\eta_\mu,$$

$$\Delta s_2=(N_{\mathbf{k}\lambda}+1)\,\eta_m\left(k-\omega_{m\mu}+i\frac{\gamma}{2}\right). \tag{34}$$

These formulas are valid provided

$$\eta_\mu=\frac{c}{(k-\omega_{m\mu})^2+\gamma^2/4}<1,$$

$$\eta_m=\frac{a}{(k-\omega_{m\mu})^2+\gamma^2/4}<1. \tag{35}$$

In the absence of excited atoms, the branch s_2 disappears and we arrive at the standard polariton spectrum, well known in the semiclassical theory of radiation. The standard polariton spectrum also follows from quantum electrodynamics.[18] Determined by the polarization operator in the technique of total cutoff of the correlators, this spectrum retains its form for excited media, requiring in this case the substitutions $n_\mu\to n_\mu-n_m$ (Ref. 2). Thus, the appearance of a third branch of the spectrum is characteristic of the photon cooperative effect. The standard polariton spectrum depends solely on the difference $n_\mu-n_m$. The sum of concentrations $n_\mu+n_m$ enters along with this difference in formulas (34). This peculiarity, like the presence of a third branch of the spectrum, owes its existence to the induced emission. It is formed by the extra photon in the virtual state of the process of interaction of an excited atom with radiation. The reason for the presence of the term $i(n_\mu+n_m)$ in the spectrum is physically obvious[13] since the state $|N_{\mathbf{k}\lambda}\rangle$ may vanish as a result of absorption of a photon in the medium, $|N_{\mathbf{k}\lambda}\rangle\to|N_{\mathbf{k}\lambda}-1\rangle$, and as a result of induced emission of a photon, $|N_{\mathbf{k}\lambda}\rangle\to|N_{\mathbf{k}\lambda}+1\rangle$. Both of the newly arising states $|N_{\mathbf{k}\lambda}\pm1\rangle$ are orthogonal to the original state.

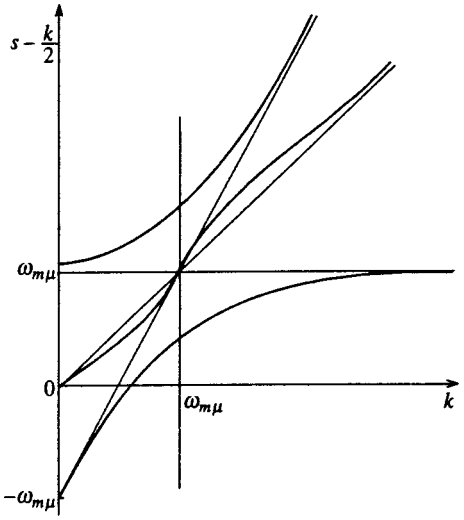


FIG. 1. Dispersion curves corresponding to the propagator Δ for $N_{\mathbf{k}\lambda}=1$.

The spectrum formed by the poles of the propagator Δ in the limit $\gamma\to0$ is plotted for $N_{\mathbf{k}\lambda}=1$ in Fig. 1.

The integral (33) in the limit $\gamma\to0$ can be calculated explicitly:

$$\Delta(\tau)=e^{-\tau s_0}\frac{(s_0-s_1^0)(s_0-s_2^{(0)})}{(s_0-s_1)(s_0-s_2)}$$
$$+e^{-\tau s_1}\frac{(s_1-s_1^{(0)})(s_1-s_2^{(0)})}{(s_1-s_0)(s_1-s_2)}$$
$$+e^{-\tau s_2}\frac{(s_2-s_1^{(0)})(s_2-s_2^{(0)})}{(s_2-s_0)(s_2-s_1)}.$$

Under conditions in which inequalities (35) are valid, guided by Eqs. (31) we find the following photon distribution over the mean filling numbers $N_{\mathbf{k}\lambda}$, which replaces the Gibbs distribution: ($\beta=1/T$):

$$Ze^{k\beta/2}\rho(N_{\mathbf{k}\lambda})$$
$$=\exp\{-\beta[kN_{\mathbf{k}\lambda}+N_{\mathbf{k}\lambda}(\eta_\mu-\eta_m)(k-\omega_{m\mu})-\eta_m$$
$$\times(k-\omega_{m\mu})]\}-\{\exp(-\beta kN_{\mathbf{k}\lambda})-\exp[-\beta(\omega_{m\mu}-k$$
$$+kN_{\mathbf{k}\lambda})]\}\eta_\mu N_{\mathbf{k}\lambda}-\{\exp(-\beta kN_{\mathbf{k}\lambda})$$
$$-\exp[-\beta(\omega_{m\mu}+k+kN_{\mathbf{k}\lambda})]\}\eta_m(1+N_{\mathbf{k}\lambda}),$$

where

$$\exp[\beta\eta_m(k-\omega_{m\mu})]Ze^{k\beta/2}$$
$$=\{1-\exp(-\beta[k+(\eta_\mu-\eta_m)(k-\omega_{m\mu})])\}$$
$$\times\left[1+\eta_\mu\frac{1-\exp[-\beta(\omega_{m\mu}-k)]}{e^{\beta k}-1}\right.$$

$$+\eta_m\frac{1-\exp[-\beta(\omega_{m\mu}-k)]}{e^{\beta k}-1}$$

$$+\eta_m(1-\exp[-\beta(k-\omega_{m\mu})])\Bigg].$$

The coefficient of η_m contains vacuum terms which disappear from the distribution function of ideal photons and which now contribute to all the moments of the distribution function. In particular,

$$\langle N_{\mathbf{k}\lambda}\rangle=\sum_{N=0}^{\infty}N\rho(N)$$

$$=\{\exp(\beta[k+(\eta_\mu-\eta_m)(k-\omega_{m\mu})])-1\}^{-1}$$

$$\times\Bigg[1+\eta_\mu\frac{1-\exp[-\beta(\omega_{m\mu}-k)]}{e^{\beta k}-1}$$

$$+\eta_m(1-\exp[-\beta(k-\omega_{m\mu})])\left(1+\frac{1}{e^{\beta k}-1}\right)\Bigg]$$

$$-\frac{\eta_\mu}{e^{\beta k}-1}\left(1+\frac{2}{e^{\beta k}-1}\right)(1-\exp[-\beta(\omega_{m\mu}-k)])$$

$$-\frac{2\eta_m}{e^{\beta k}-1}\left(1+\frac{1}{e^{\beta k}-1}\right)(1-\exp[-\beta(k-\omega_{m\mu})]). \tag{36}$$

We have arrived here at a generalization of the Planck formula in terms of wavelengths. The generalizations of this formula in terms of wavelengths and in terms of frequencies are fundamentally different.[19] The ''overgrowth'' of the mean values of the filling numbers in the Green's function technique based on cutting off the correlators was investigated in Ref. 20. The results of this work in our notation for $\eta_{\mu,m}<1$ looks like this:

$$\langle N_{\mathbf{k}\lambda}\rangle=\{\exp(\beta[k+(\eta_\mu-\eta_m)(k-\omega_{m\mu})])-1\}^{-1}$$

$$-\frac{\eta_\mu-\eta_m}{e^{\beta k}-1}+\frac{\eta_\mu-\eta_m}{e^{\beta\omega_{m\mu}}-1}. \tag{37}$$

As already noted, in a technique of this kind corrections for nonideal behavior depend only on the difference $\eta_\mu-\eta_m$, but not on either of these quantities separately. The account of cooperative phenomena contained in formula (36) is substantially different from that contained in formula (37). This difference, as could be expected, is not analytical in the charge. For $k/T\gg 1$ the role of induced processes in the formation of the equilibrium distribution is small. Under these conditions, formulas (36) and (37) coincide. Here it is necessary to bear in mind that the presence of any appreciable values of

$$\eta_m\sim\chi^3 n_m\frac{\gamma_r\omega_{m\mu}}{(k-\omega_{m\mu})^2+\gamma^2/4}$$

requires that $k\to\omega_{m\mu}$.

11. CONCLUSION

The photon cooperative effect, closely associated with the induced emission of atoms, is manifested quantitatively for $\eta_{\mu,m}\sim 1$, i.e., under conditions near resonance. This does not mean that the effect is important only for frequencies inside the spectral contour of the spontaneous emission of the atoms of the medium. As follows from the difference between formulas (36) and (37), it is also manifested when $|\omega_{m\mu}-k|>\gamma$. For $|\omega_{m\mu}-k|<\gamma$ the role of this effect is enhanced thanks to the appearance of the term $i\gamma(n_\mu+n_m)$. The standard refractive index of the medium, which falls back to unity if $n_\mu=n_m$, cannot depend on $n_{\mu,m}$ in this way.

The occurrence of a cooperative effect under equilibrium conditions obviously implies its occurrence in kinetic phenomena. In this context its experimental confirmation is significantly simpler, and the appearance of the effect itself may be expected on the macroscopic level. In nonequilibrium situations, the incoherent reaction channel, as was noted above, does not admit a perturbation-theory treatment and must be taken into account. The coherent and incoherent channels must be studied differently since, as a result of orthogonality of the wave functions of the medium in the final states these channels do not interfere. Besides, in a number of situations the incoherent channel can be shown[13] to be positive definite. Under such conditions, a study of just the coherent channel makes it possible to obtain a lower estimate for a number of phenomena. Precisely in this way it was shown in Ref. 13 that the coefficient of reflection R of resonant radiation from a thermally excited medium is given by $R\propto|i\gamma(n_\mu+n_m)|^2$, which is characteristic for a cooperative effect. In contrast to the predictions of semiclassical theory, $R\neq 0$ for $n_\mu=n_m$. In the standard technique of quantum Green's functions[1] which presupposes a cutoff in all the correlators, such a result cannot be obtained. The manifestation of a cooperative effect can also be expected in other optical phenomena. It is precisely this effect, and not the standard refractive index, that determines the intensity of the transient back-emission[21] in excited media in the resonant frequency region. It may be supposed that in the description of all coherent optical phenomena in excited media in the quasiresonant region the standard refractive index must be used with care. Indeed, the standard refractive index is algebraically related to the polarization operator $\hat{\Pi}_{\mu\nu}$ arising in the Green's function technique,[1] which excludes cooperative effects. An account of the latter requires the introduction of the polarization operator $\hat{\mathscr{P}}$ in the Γ-operator technique. In the quasiresonant region the operators $\hat{\Pi}_{\mu\nu}$ and $\hat{\mathscr{P}}$ do not coincide. Moreover, these operators generate different polariton spectra for $n_m\neq 0$. The appearance of a specific dispersion relation due to the operator $\hat{\mathscr{P}}$ and the appearance of a third branch of the spectrum raises the question of their manifestations in experimentally realizable situations. The dispersion relation by itself governs Vavilov–Čerenkov radiation, transient forward-emission, and light refraction phenomena at interfaces. For $n_m\neq 0$ peculiarities may be expected in these phenomena. Conditions facilitating the

appearance of such peculiarities for thermally excited media are $k \to \omega_{m\mu}$ and $n_m \to n_\mu$.

*)E-mail: phf@deans.mpei.ac.ru

[1] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, N.J., 1963).
[2] A. A. Panteleev, V. A. Roslyakov, and A. N. Starostin, Zh. Éksp. Teor. Fiz. **97**, 1777 (1990) [Sov. Phys. JETP **70**, 1003 (1990)].
[3] Yu. K. Zemtsov, A. Yu. Sechin, and A. N. Starostin, Zh. Éksp. Teor. Fiz. **110**, 1654 (1996) [JETP **83**, 909 (1996)].
[4] R. H. Dicke, Phys. Rev. **93**, 99 (1954).
[5] M. J. Stephen, J. Chem. Phys. **40**, 669 (1964).
[6] R. Bonifacio and P. Schwendimann, Phys. Rev. A **4**, 302, 854 (1971).
[7] F. Haake and R. J. Glauber, Phys. Rev. A **5**, 1457 (1972).
[8] M. S. Feld and J. C. MacGillivary, in *Topics in Current Physics*, Vol. 21, edited by M. S. Feld and V. S. Letokhov (Springer, Berlin, 1980), p. 7.
[9] G. S. Agarwal and R. R. Puri, Opt. Commun. **69**, 267 (1989).
[10] G. M. Palma, A. Vaglica, C. Leonardi, *et al.*, Opt. Commun. **377**, 79 (1990).
[11] M. R. Wahiddin, S. S. Hassan, and R. K. Bullough, J. Mod. Opt. **42**, 171 (1995).
[12] B. A. Veklenko, Izv. Vuzov SSSR. Fizika, No. 9, 71 (1983).
[13] B. A. Veklenko, Zh. Éksp. Teor. Fiz. **96**, 457 (1989) [Sov. Phys. JETP **69**, 258 (1989)].
[14] B. A. Veklenko and G. B. Tkachuk, Izv. Vuzov. Fizika, No. 2, 89 (1987).
[15] M. J. Gagen, H. M. Viseman, and G. J. Milburn, Phys. Rev. A **48**, 132 (1993).
[16] M. B. Mensky, Phys. Lett. A **219**, 137 (1996).
[17] A. I. Akhiezer and V. B. Berestetskii, *Quantum Electrodynamics* (Wiley, New York, 1965).
[18] Yu. A. Vdovin and V. M. Galitskiĭ, Zh. Éksp. Teor. Fiz. **48**, 1352 (1965) [Sov. Phys. JETP **21**, 904 (1965)].
[19] V. L. Ginzburg and L. P. Pitaevskiĭ, Usp. Fiz. Nauk **151**, 333 (1987) [Sov. Phys. Usp. **30**, 168 (1987)].
[20] N. N. Bogolyubov and S. V. Tyablikov, Dokl. Nauk SSSR **126**, 53 (1959) [Sov. Phys. Dokl. **4**, 589 (1959)].
[21] B. A. Veklenko, Ivz. Vuzov. Fizika, No. 11, 11 (1984).

Translated by Paul F. Schippnick

Optical pumping of velocity-selective coherent population trapping states in three-level atoms

V. G. Minogin*)

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia

G. Nienhuis

Huygens Laboratorium, Rijksuniversiteit Leiden, Postbus 9504 2300 RA Leiden, The Netherlands
(Submitted 30 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 511–525 (August 1998)

An analytical theory of optical pumping of velocity-selective coherent population trapping states in three-level Λ-atoms is presented. The theory is based on an atomic density matrix in the superpositional state representation. The representation introduced for the atomic density matrix directly describes coherent population trapping (CPT) states in a three-level Λ-atom and yields an approximate analytical description for the shape of the two-peak momentum (velocity) distribution of atoms. Optical pumping of velocity-selective CPT states is described in the superpositional state representation as a process of redistribution of atoms in the region of small velocities due to the photon recoil. Typical times of formation of velocity-selective CPT states are found to be inversely proportional to the square of the recoil energy. © *1998 American Institute of Physics.* [S1063-7761(98)01108-1]

1. INTRODUCTION

The interaction of multilevel atoms with optical fields is essentially different from that of two-level atoms because of coherence effects due to interference between the different excitation channels. The best known of these effects is the coherent population trapping (CPT) effect in three-level atoms with the so-called Λ-scheme of energy levels.[1,2] At certain detunings between two optical waves exciting an atom on two adjacent transitions, a Λ-atom can be driven to a superpositional CPT state composed of the two lower states. In such a state, the atom is optically decoupled from the upper state owing to destructive interference between two excitation channels. The CPT effect is a fairly common feature of multilevel systems and occurs in all cases when conditions for the destructive interference between excitation channels are satisfied.[3–5] The influence of the CPT effect on the dynamics of internal transitions in multilevel atoms has been studied experimentally.[2,6,7] In recent years, the CPT effect has attracted a lot of attention in connection with its use for formation of velocity-selective CPT states in laser-cooled atomic ensembles.[8–12]

From the theoretical standpoint, one of the basic problems in the analysis of velocity-selective superpositional states of atoms is the analytical description of formation of CPT states. A solution to this problem is needed both for an adequate description of time-dependent momentum distributions of multilevel atoms in multifrequency optical fields and for analytical estimates of formation rates of CPT states. In the standard approach, the formation of CPT states is usually described by equations for the atomic density matrix $\rho_{\alpha\beta}$ in the basis of unperturbed bare atomic states $|\alpha\rangle$ and $\beta\rangle$.

This paper presents an analysis of formation of velocity-selective CPT states based on a description of the atomic dynamics in terms of the density matrix in the superpositional state representation. This representation has considerable advantages over the conventional representation since it directly describes dynamics of superpositional states coupled with and decoupled from an external optical field. This representation was previously used in part for qualitative estimates of the momentum (velocity) width of the atomic distribution produced by a velocity-selective CPT effect.[9] A general theoretical description of this approach was discussed by Kocharovskaya *et al.*[13] For definiteness, in this paper, we analyze the simplest (but at the same time practically important) case of a three-level Λ-atom with two ground-state sublevels and one optically excited upper level. We assume that a Λ-atom interacts with two plane optical waves each of which is considered to be at resonance with an optical transition from one of the ground-state sublevel to the upper state.

2. SUPERPOSITIONAL STATE REPRESENTATION FOR PROBABILITY AMPLITUDES

Consider first the basic equations describing the coherent dynamics of atomic wave packet of a three-level atom interacting with an optical field composed of two traveling light waves:

$$\mathbf{E}=\mathbf{E}_1\cos(\mathbf{k}_1\cdot\mathbf{r}-\omega_1 t)+\mathbf{E}_2\cos(\mathbf{k}_2\cdot\mathbf{r}-\omega_2 t), \tag{1}$$

where the first wave with frequency ω_1 is assumed to be close to resonance with the $|1\rangle\rightarrow|3\rangle$ atomic transition, and the second is close to the $|2\rangle\rightarrow|3\rangle$ transition (Fig. 1a).

If the interaction times are shorter than the spontaneous decay time of the upper level, the evolution of the atomic wave packet can be described by the Schrödinger equation with the Hamiltonian

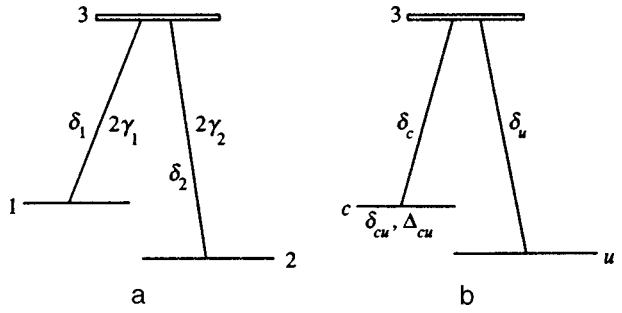


FIG. 1. Three-level Λ-atom (a) in the standard representation and (b) in the superpositional state representation.

$$H=-(\hbar^2/2M)\Delta+H_0-\mathbf{D}\cdot\mathbf{E}, \tag{2}$$

where the Hamiltonian H_0 determines the states of an unperturbed atom, $V=-\mathbf{D}\cdot\mathbf{E}$ is the dipole interaction operator, and $\mathbf{D}$ is the dipole moment operator. The total wave function in the coordinate representation describing the internal and translational motion of a Λ-atom can be expressed as usual through atomic probability amplitudes $a=a(\mathbf{r},t)$:

$$\Psi(\mathbf{r},\xi,t)=\sum a_\alpha(\mathbf{r},t)\psi_\alpha(\xi)\exp\left(-\frac{i\varepsilon_\alpha t}{\hbar}\right), \tag{3}$$

where $\psi_\alpha(\xi)$ are unperturbed wave functions describing the motion of electrons in the atom in states $|\alpha\rangle$, $\alpha=3,2,1$. The probability amplitudes $a_\alpha=a_\alpha(\mathbf{r},t)$ satisfy equations which include explicit coordinate and time dependence:

$$\begin{aligned}
i\frac{\partial}{\partial t}a_3&=-\frac{\hbar}{2M}\Delta a_3-g_1\exp[i(\mathbf{k}_1\cdot\mathbf{r}-\delta_1 t)]a_1-g_2\\
&\quad\times\exp[i(\mathbf{k}_2\cdot\mathbf{r}-\delta_2 t)]a_2,\\
i\frac{\partial}{\partial t}a_2&=-\frac{\hbar}{2M}\Delta a_2-g_2\exp[-i(\mathbf{k}_2\cdot\mathbf{r}-\delta_2 t)]a_3,\\
i\frac{\partial}{\partial t}a_1&=-\frac{\hbar}{2M}\Delta a_1-g_1\exp[-i(\mathbf{k}_1\cdot\mathbf{r}-\delta_1 t)]a_3,
\end{aligned} \tag{4}$$

where g_1 and g_2 are equal to half the Rabi frequencies, δ_1 and δ_2 are the detuning parameters:

$$g_1=\mathbf{d}_{13}\cdot\mathbf{E}_1/2\hbar, \quad g_2=\mathbf{d}_{23}\cdot\mathbf{E}_2\hbar, \tag{5}$$

$$\delta_1=\omega_1-\omega_{31}, \quad \delta_2=\omega_2-\omega_{32}, \tag{6}$$

and $\omega_{\alpha\beta}=(\varepsilon_\alpha-\varepsilon_\beta)/\hbar$ are the frequencies of atomic transitions. For simplicity, the matrix elements of the atomic dipole moments, $\mathbf{d}_{\alpha\beta}$, are assumed to be real.

In analyzing the coherent dynamics of a Λ-atom determined by Eqs. (4), it is natural to introduce new amplitudes $b_\alpha(\mathbf{r},t)$ that satisfy equations without an explicit dependence on space and time. Given Eqs. (4), we define the new probability amplitudes by the relations

$$\begin{aligned}
&a_3=b_3, \quad a_2=b_2\exp[-i(\mathbf{k}_2\cdot\mathbf{r}-\delta_2 t)],\\
&a_1=b_1\exp[-i(\mathbf{k}_1\cdot\mathbf{r}-\delta_1 t)].
\end{aligned} \tag{7}$$

The new amplitudes $b_\alpha=b_\alpha(\mathbf{r},t)$, $\alpha=3,2,1$, satisfy equations that contain only an implicit space and time dependence:

$$\begin{aligned}
i\frac{\partial}{\partial t}b_3&=-\frac{\hbar}{2M}\Delta b_3-g_1b_1-g_2b_2,\\
i\frac{\partial}{\partial t}b_2&=-\frac{\hbar}{2M}\Delta b_2-g_2b_3+(\delta_2+\omega_r)b_2+i\mathbf{v}_2\nabla b_2,\\
i\frac{\partial}{\partial t}b_1&=-\frac{\hbar}{2M}\Delta b_1-g_1b_3+(\delta_1+\omega_r)b_1+i\mathbf{v}_1\nabla b_1,
\end{aligned} \tag{8}$$

where $\nabla=\partial/\partial\mathbf{r}$, and $\mathbf{v}_1=\hbar\mathbf{k}_1/M$ and $\mathbf{v}_2=\hbar\mathbf{k}_2/M$ are the recoil velocities. In the present case of two light waves of close frequencies, $\omega_1\simeq\omega_2$, and wave vectors $|\mathbf{k}_1|\simeq|\mathbf{k}_2|\simeq k$, the absolute values of recoil velocities are approximately equal: $|\mathbf{v}_1|\simeq|\mathbf{v}_2|\simeq v_r=\hbar k/M$. The frequency ω_r determines the total recoil energy $R=\hbar\omega_r$:

$$\omega_r=\hbar k^2/2M. \tag{9}$$

The character of the coherent dynamics of a three-level Λ-atom in an electromagnetic field (1) is clearly seen from the equations for probability amplitudes $b_\alpha=b_\alpha(\mathbf{p},t)$ in the momentum representation:

$$\begin{aligned}
i\frac{\partial}{\partial t}b_3&=\frac{\mathbf{p}^2}{2M\hbar}b_3-g_1b_1-g_2b_2,\\
i\frac{\partial}{\partial t}b_2&=\frac{\mathbf{p}^2}{2M\hbar}b_2-g_2b_3+[\Delta(\mathbf{v})-\delta(\mathbf{v})]b_2,\\
i\frac{\partial}{\partial t}b_1&=\frac{\mathbf{p}^2}{2M\hbar}b_1-g_1b_3+[\Delta(\mathbf{v})+\delta(\mathbf{v})]b_1,
\end{aligned} \tag{10}$$

where $\mathbf{v}=\mathbf{p}/M$ is the atom velocity, $\Delta(\mathbf{v})$ and $\delta(\mathbf{v})$ are the velocity-dependent detunings:

$$\begin{aligned}
\Delta(\mathbf{v})&=\frac{1}{2}[(\delta_1-\mathbf{k}_1\cdot\mathbf{v})+(\delta_2-\mathbf{k}_2\cdot\mathbf{v})]+\omega_r,\\
\delta(\mathbf{v})&=\frac{1}{2}[(\delta_1-\mathbf{k}_1\cdot\mathbf{v})-(\delta_2-\mathbf{k}_2\cdot\mathbf{v})].
\end{aligned} \tag{11}$$

Note that in the momentum representation Eq. (7) transforms to

$$\begin{aligned}
&a_3(\mathbf{p})=b_3(\mathbf{p}), \quad a_2(\mathbf{p})=e^{i\delta_2 t}b_2(\mathbf{p}+\hbar\mathbf{k}_2),\\
&a_1(\mathbf{p})=e^{i\delta_1 t}b_1(\mathbf{p}+\hbar\mathbf{k}_1).
\end{aligned} \tag{12}$$

The last two lines of Eqs. (10) clearly show that, under the condition of coherent population trapping,

$$\delta(\mathbf{v})=0, \tag{13}$$

there is always a specific CPT state, in which the atomic wave packet is decoupled from the exciting optical field. Given the structure of Eqs. (11), the existence of a CPT state becomes obvious if, keeping the amplitude of the upper state unchanged, one introduces superpositional probability amplitudes $b_c(\mathbf{p},t)$ and $b_u(\mathbf{p},t)$ for the ground atomic states[1,2]:

$$b_c=b_1\sin\theta+b_2\cos\theta, \quad b_u=b_1\cos\theta-b_2\sin\theta, \tag{14}$$

where

$$\sin\theta=g_1/g_0, \quad \cos\theta=g_2/g_0, \quad g_0^2=g_1^2+g_2^2, \tag{15}$$

and the subscripts c and u refer to the lower superposition states optically coupled with and decoupled from, respectively, the upper Λ-atom state under the CPT condition (13).

Equations for the probability amplitude $b_3=b_3(\mathbf{p},t)$ and superpositional probability amplitudes $b_c=b_c(\mathbf{p},t)$ and $b_u=b_u(\mathbf{p},t)$, in accordance with definition (14) and Eqs. (10), are

$$\begin{aligned}
i\frac{\partial}{\partial t}b_3&=\frac{\mathbf{p}^2}{2M\hbar}b_3-g_0b_c,\\
i\frac{\partial}{\partial t}b_c&=\frac{\mathbf{p}^2}{2M\hbar}b_c-g_0b_3+[\Delta(\mathbf{v})-\delta(\mathbf{v})\cos 2\theta]b_c\\
&\quad+\delta(\mathbf{v})b_u\sin 2\theta,\\
i\frac{\partial}{\partial t}b_u&=\frac{\mathbf{p}^2}{2M\hbar}b_u+[\Delta(\mathbf{v})+\delta(\mathbf{v})\cos 2\theta]b_u\\
&\quad+\delta(\mathbf{v})b_c\sin 2\theta.
\end{aligned}\tag{16}$$

The last line of Eqs. (16) clearly shows that, under condition (13), the superpositional state $b_u=b_u(\mathbf{p},t)$ is decoupled from other atomic states. Thus, the representation of superpositional states (14) can be considered as a natural representation for describing the coherent population trapping effect in a Λ-atom.

In the coordinate representation, the structure of Eq. (14) is basically the same. In this case, the equations for the superpositional probability amplitudes $b_\alpha=b_\alpha(\mathbf{r},t)$, $\alpha=3,c,u$ have the form

$$\begin{aligned}
i\frac{\partial}{\partial t}b&=-\frac{\hbar}{2M}\Delta b_3-g_0b_c,\\
i\frac{\partial}{\partial t}b_c&=-\frac{\hbar}{2M}\Delta b_c-g_0b_3+(\delta_1\sin^2\theta+\delta_2\cos^2\theta+\omega_r)b_c\\
&\quad+\frac{1}{2}(\delta_1-\delta_2)b_u\sin 2\theta+i(\mathbf{v}_1\sin^2\theta+\mathbf{v}_2\cos^2\theta)\\
&\quad\times\nabla b_c+\frac{i}{2}(\mathbf{v}_1-\mathbf{v}_2)\nabla b_u\sin 2\theta,\\
i\frac{\partial}{\partial t}b_u&=-\frac{\hbar}{2M}\Delta b_u+\frac{1}{2}(\delta_1-\delta_2)b_c\sin 2\theta\\
&\quad+(\delta_1\cos^2\theta+\delta_2\sin^2\theta+\omega_r)b_u+\frac{i}{2}\\
&\quad\times(\mathbf{v}_1-\mathbf{v}_2)\nabla b_c\sin 2\theta+i(\mathbf{v}_1\cos^2\theta+\mathbf{v}_2\sin^2\theta)\nabla b_u.
\end{aligned}\tag{17}$$

Physically, the set of equations (17) is equivalent to Eqs. (8). Mathematically, the difference between Eqs. (8) and (17) is caused by the use of different representations, namely, the standard representation of bare atomic states $|2\rangle$ and $|1\rangle$ in Eqs. (8), and the superpositional state $|c\rangle$ and $|u\rangle$ representation in Eqs. (17).

Note that the normalization condition for the superpositional probability amplitudes b_α, $\alpha=3,c,u$, in both the coordinate and momentum representations has the usual form:

$$\int[|b_c|^2+|b_u|^2+|b_3|^2]d^3r=1,\tag{18a}$$

$$\int[|b_c|^2+|b_u|^2+|b_3|^2]d^3p=1.\tag{18b}$$

3. SUPERPOSITIONAL STATE REPRESENTATION FOR ATOMIC DENSITY MATRIX

The superpositional state representation introduced above allows us to generalize the description of the CPT effect in a Λ-atom by taking into account spontaneous relaxation.

A natural approach to the description of quantum statistical states of an atomic wave packet in a resonant light field is based on an atomic density matrix in the Wigner representation, $\rho=\rho(\mathbf{r},\mathbf{p},t)$.[14] The relation between the Wigner representation and the coordinate representation of the density matrix is described by the well-known equations

$$\begin{aligned}
\rho(\mathbf{r},\mathbf{p})&=\frac{1}{(2\pi\hbar)^3}\int\rho\left(\mathbf{r}+\frac{\mathbf{s}}{2},\mathbf{r}-\frac{\mathbf{s}}{2}\right)\exp\left(-\frac{i\mathbf{p}\cdot\mathbf{s}}{\hbar}\right)d^3s,\\
\rho(\mathbf{r},\mathbf{r}')&=\int\rho\left(\frac{\mathbf{r}+\mathbf{r}'}{2},\mathbf{p}\right)\exp\left[\frac{i(\mathbf{r}-\mathbf{r}')\cdot\mathbf{p}}{\hbar}\right]d^3p.
\end{aligned}\tag{19}$$

The density matrix ρ is expressed in terms of the probability amplitudes a_α, $\alpha=3,2,1$. In the case of pure atomic states in the coordinate representation,

$$\rho_{\alpha\beta}(\mathbf{r},\mathbf{r}')=a_\alpha(\mathbf{r})a_\beta^*(\mathbf{r}').\tag{20}$$

The matrix elements of the Wigner density matrix $\rho(\mathbf{r},\mathbf{p},t)$, which describes the dynamics of a Λ-atom in the field (1), satisfy the equations

$$\begin{aligned}
\frac{d}{dt}\rho_{33}(\mathbf{p})&=ig_1\exp[i(\mathbf{k}_1\cdot\mathbf{r}-\delta_1t)]\rho_{13}\left(\mathbf{p}-\frac{\hbar\mathbf{k}_1}{2}\right)\\
&\quad+ig_2\exp[i(\mathbf{k}_2\cdot\mathbf{r}-\delta_2t)]\rho_{23}\left(\mathbf{p}-\frac{\hbar\mathbf{k}_2}{2}\right)\\
&\quad+\text{c.c.}-2(\gamma_1+\gamma_2)\rho_{33}(\mathbf{p}),\\
\frac{d}{dt}\rho_{22}(\mathbf{p})&=ig_2\exp[-i(\mathbf{k}_2\cdot\mathbf{r}-\delta_2t)]\rho_{32}\left(\mathbf{p}+\frac{\hbar\mathbf{k}_2}{2}\right)\\
&\quad+\text{c.c.}+2\gamma_2\int\Phi(\hat{\mathbf{n}})\rho_{33}(\mathbf{p}+\hat{\mathbf{n}}\hbar k)d\hat{\mathbf{n}},\\
\frac{d}{dt}\rho_{11}(\mathbf{p})&=ig_1\exp[-i(\mathbf{k}_1\cdot\mathbf{r}-\delta_1t)]\rho_{31}\left(\mathbf{p}+\frac{\hbar\mathbf{k}_1}{2}\right)\\
&\quad+\text{c.c.}+2\gamma_1\int\Phi(\hat{\mathbf{n}})\rho_{33}(\mathbf{p}+\hat{\mathbf{n}}\hbar k)d\hat{\mathbf{n}},\\
\frac{d}{dt}\rho_{31}(\mathbf{p})&=ig_1\exp[-i(\mathbf{k}_1\cdot\mathbf{r}-\delta_1t)]\left[\rho_{11}\left(\mathbf{p}-\frac{\hbar\mathbf{k}_1}{2}\right)\right.\\
&\quad\left.-\rho_{33}\left(\mathbf{p}+\frac{\hbar\mathbf{k}_1}{2}\right)\right]+ig_2\exp[i(\mathbf{k}_2\cdot\mathbf{r}-\delta_2t)]\\
&\quad\times\rho_{21}\left(\mathbf{p}-\frac{\hbar\mathbf{k}_2}{2}\right)-(\gamma_1+\gamma_2)\rho_{31}(\mathbf{p}),
\end{aligned}$$

$$\frac{d}{dt}\rho_{32}(\mathbf{p}) = ig_2 \exp[i(\mathbf{k}_2\cdot\mathbf{r}-\delta_2 t)]\left[\rho_{22}\left(\mathbf{p}-\frac{\hbar\mathbf{k}_2}{2}\right)-\rho_{33}\times\left(\mathbf{p}+\frac{\hbar\mathbf{k}_2}{2}\right)\right]+ig_1\exp[i(\mathbf{k}_1\cdot\mathbf{r}-\delta_1 t)]\rho_{12}\times\left(\mathbf{p}-\frac{\hbar\mathbf{k}_1}{2}\right)-(\gamma_1+\gamma_2)\rho_{32}(\mathbf{p}),$$

$$\frac{d}{dt}\rho_{12}(\mathbf{p}) = ig_1\exp[-i(\mathbf{k}_1\cdot\mathbf{r}-\delta_1 t)]\rho_{32}\left(\mathbf{p}+\frac{\hbar\mathbf{k}_1}{2}\right)-ig_2\exp[i(\mathbf{k}_2\cdot\mathbf{r}-\delta_2 t)]\rho_{13}\left(\mathbf{p}+\frac{\hbar\mathbf{k}_2}{2}\right), \quad (21)$$

where $\rho(\mathbf{p})=\rho(\mathbf{r},\mathbf{p},t)$, d/dt denotes the total (convective) time derivative

$$d/dt = \partial/\partial t + \mathbf{v}\cdot\partial/\partial\mathbf{r}, \quad (22)$$

and $\hat{\mathbf{n}}$ is the unit vector which defines the direction of a photon emitted in a single spontaneous decay act. The function $\Phi(\hat{\mathbf{n}})$ describes the spatial asymmetry of photon emission, and hence the anisotropy of the photon recoil. The specific form of $\Phi(\hat{\mathbf{n}})$ is determined by the polarization of light waves.[14] This function is normalized to unity:

$$\int \Phi(\hat{\mathbf{n}})d\hat{\mathbf{n}} = 1. \quad (23)$$

The elements of the density matrix satisfy the conventional Hermitian condition $\rho^*_{\alpha\beta}(\mathbf{r},\mathbf{p},t)=\rho_{\beta\alpha}(\mathbf{r},\mathbf{p},t)$. The Rabi frequencies and detunings in Eqs. (21) are determined by Eqs. (5) and (6).

The Wigner function $W(\mathbf{r},\mathbf{p},t)$, which describes the shape of the atomic wave packet, is defined by the equation

$$W(\mathbf{r},\mathbf{p},t)=\rho_{11}(\mathbf{r},\mathbf{p},t)+\rho_{22}(\mathbf{r},\mathbf{p},t)+\rho_{33}(\mathbf{r},\mathbf{p},t). \quad (24)$$

The Wigner function is assumed to be normalized to unity for a single atom:

$$\int W(\mathbf{r},\mathbf{p},t)d^3r\,d^3p = 1. \quad (25)$$

Equations (21) can be transformed into the superpositional state representation using the procedure based on the relation between the elements of the density matrix and probability amplitudes for pure atomic states. In order to perform this procedure, we introduce a density matrix σ related to the probability amplitudes b_α. For pure atomic states, the density matrix elements $\sigma_{\alpha\beta}$ are related to the amplitudes b_α by equations having the form of Eq. (20):

$$\sigma_{\alpha\beta}(\mathbf{r},\mathbf{r}') = b_\alpha(\mathbf{r})b^*_\beta(\mathbf{r}'), \quad (26)$$

where $\alpha,\beta=3,2,1$ in the standard representation of bare atomic states and $\alpha,\beta=3,c,u$ in the superpositional state representation.

The Wigner representation and coordinate representation for density matrix σ are related by formulas similar to Eqs. (19) after the change $\rho\to\sigma$. Using these relations and Eqs. (7), (20) and (26), it is convenient to determine first the correspondence between the Wigner density matrices $\rho(\mathbf{r},\mathbf{p},t)=\rho(\mathbf{p})$ and $\sigma(\mathbf{r},\mathbf{p},t)=\sigma(\mathbf{p})$ in the standard representation of bare atomic states:

$$\sigma_{33}(\mathbf{p})=\rho_{33}(\mathbf{p}), \quad \sigma_{22}(\mathbf{p})=\rho_{22}(\mathbf{p}-\hbar\mathbf{k}_2),$$
$$\sigma_{11}(\mathbf{p})=\rho_{11}(\mathbf{p}-\hbar\mathbf{k}_1),$$
$$\sigma_{31}(\mathbf{p})=\exp[-i(\mathbf{k}_1\cdot\mathbf{r}-\delta_1 t)]\rho_{31}\left(\mathbf{p}-\frac{\hbar\mathbf{k}_1}{2}\right),$$
$$\sigma_{32}(\mathbf{p})=\exp[-i(\mathbf{k}_2\cdot\mathbf{r}-\delta_2 t)]\rho_{32}\left(\mathbf{p}-\frac{\hbar\mathbf{k}_2}{2}\right),$$
$$\sigma_{12}(\mathbf{p})=\exp[i(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{r}-i(\delta_1-\delta_2)t]\times\rho_{12}\left(\mathbf{p}-\frac{\hbar(\mathbf{k}_1+\mathbf{k}_2)}{2}\right). \quad (27)$$

Using once again the relation between the Wigner and coordinate representations and taking into account Eqs. (14), (20), and (26), one can find the relation between elements of the Wigner density matrix σ in the standard representation of bare atomic states and in the superpositional state representation:

$$\sigma_{33}=\sigma_{33},$$
$$\sigma_{cc}=\sigma_{11}\sin^2\theta+\sigma_{22}\cos^2\theta+(\sigma_{12}+\sigma_{21})\sin\theta\cos\theta,$$
$$\sigma_{uu}=\sigma_{11}\cos^2\theta+\sigma_{22}\sin^2\theta-(\sigma_{12}+\sigma_{21})\sin\theta\cos\theta,$$
$$\sigma_{3c}=\sigma_{31}\sin\theta+\sigma_{32}\cos\theta, \quad \sigma_{3u}=\sigma_{31}\cos\theta-\sigma_{32}\sin\theta,$$
$$\sigma_{cu}=(\sigma_{11}-\sigma_{22})\sin\theta\cos\theta+\sigma_{21}\cos^2\theta-\sigma_{12}\sin^2\theta. \quad (28a)$$

The formulas for the inverse transform of elements of the Wigner density matrix σ can be derived from the above equations:

$$\sigma_{22}=\sigma_{cc}\cos^2\theta+\sigma_{uu}\sin^2\theta-(\sigma_{cu}+\sigma_{uc})\sin\theta\cos\theta,$$
$$\sigma_{11}=\sigma_{cc}\sin^2\theta+\sigma_{uu}\cos^2\theta+(\sigma_{cu}+\sigma_{uc})\sin\theta\cos\theta,$$
$$\sigma_{31}=\sigma_{3c}\sin\theta+\sigma_{3u}\cos\theta, \quad \sigma_{32}=\sigma_{3c}\cos\theta-\sigma_{3u}\sin\theta,$$
$$\sigma_{12}=(\sigma_{cc}-\sigma_{uu})\sin\theta\cos\theta-\sigma_{cu}\sin^2\theta+\sigma_{uc}\cos^2\theta. \quad (28b)$$

Finally, by calculating total time derivatives on the left-hand and right-hand sides in Eqs. (28a) and using Eqs. (21), (27), and (28a), (28b), we obtain the following equations for the Wigner density matrix $\sigma(\mathbf{r},\mathbf{p},t)=\sigma(\mathbf{p})$ in the superpositional state representation:

$$\frac{d}{dt}\sigma_{33}(\mathbf{p}) = ig_0[\sigma_{c3}(\mathbf{p})-\sigma_{3c}(\mathbf{p})]-2(\gamma_1+\gamma_2)\sigma_{33}(\mathbf{p}),$$

$$\frac{d}{dt}\sigma_{cc}(\mathbf{p}) = ig_0[\sigma_{3c}(\mathbf{p})-\sigma_{c3}(\mathbf{p})]+i\delta_{cu}(\mathbf{v})[\sigma_{cu}(\mathbf{p})-\sigma_{uc}(\mathbf{p})]+2\int\Phi(\hat{\mathbf{n}})[\gamma_1\sigma_{33}(\mathbf{p}-\hbar\mathbf{k}_1+\hat{\mathbf{n}}\hbar k)\times\sin^2\theta+\gamma_2\sigma_{33}(\mathbf{p}-\hbar\mathbf{k}_2+\hat{\mathbf{n}}\hbar k)\cos^2\theta]d\hat{\mathbf{n}}$$

$$+(\mathbf{v}_1\sin^2\theta+\mathbf{v}_2\cos^2\theta)\frac{\partial}{\partial\mathbf{r}}\sigma_{cc}(\mathbf{p})$$

$$+\frac{1}{4}(\mathbf{v}_1-\mathbf{v}_2)\sin 2\theta\frac{\partial}{\partial\mathbf{r}}[\sigma_{cu}(\mathbf{p})+\sigma_{uc}(\mathbf{p})],$$

$$\frac{d}{dt}\sigma_{uu}(\mathbf{p})=-i\delta_{cu}(\mathbf{v})[\sigma_{cu}(\mathbf{p})-\sigma_{uc}(\mathbf{p})]+2\int\Phi(\hat{\mathbf{n}})$$

$$\times[\gamma_1\sigma_{33}(\mathbf{p}-\hbar\mathbf{k}_1+\hat{\mathbf{n}}\hbar k)\cos^2\theta+\gamma_2\sigma_{33}$$

$$\times(\mathbf{p}-\hbar\mathbf{k}_2+\hat{\mathbf{n}}\hbar k)\sin^2\theta]d\hat{\mathbf{n}}$$

$$+(\mathbf{v}_1\cos^2\theta+\mathbf{v}_2\sin^2\theta)\frac{\partial}{\partial\mathbf{r}}\sigma_{uu}(\mathbf{p})$$

$$+\frac{1}{4}(\mathbf{v}_1-\mathbf{v}_2)\sin 2\theta\frac{\partial}{\partial\mathbf{r}}[\sigma_{cu}(\mathbf{p})+\sigma_{uc}(\mathbf{p})],$$

$$\frac{d}{dt}\sigma_{3c}(\mathbf{p})=ig_0[\sigma_{cc}(\mathbf{p})-\sigma_{33}(\mathbf{p})]+i\delta_{cu}(\mathbf{v})\sigma_{3u}(\mathbf{p})$$

$$+[i\delta_c(\mathbf{v})-(\gamma_1+\gamma_2)]\sigma_{3c}(\mathbf{p})$$

$$+\frac{1}{2}(\mathbf{v}_1\sin^2\theta+\mathbf{v}_2\cos^2\theta)\frac{\partial}{\partial\mathbf{r}}\sigma_{3c}(\mathbf{p})$$

$$+\frac{1}{4}(\mathbf{v}_1-\mathbf{v}_2)\sin 2\theta\frac{\partial}{\partial\mathbf{r}}\sigma_{3u}(\mathbf{p}),$$

$$\frac{d}{dt}\sigma_{3u}(\mathbf{p})=ig_0\sigma_{cu}(\mathbf{p})+[i\delta_u(\mathbf{v})-(\gamma_1+\gamma_2)]\sigma_{3u}(\mathbf{p})$$

$$+i\delta_{cu}(\mathbf{v})\sigma_{3c}(\mathbf{p})+\frac{1}{2}(\mathbf{v}_1\cos^2\theta+\mathbf{v}_2\sin^2\theta)$$

$$\times\frac{\partial}{\partial\mathbf{r}}\sigma_{3u}(\mathbf{p})+\frac{1}{4}(\mathbf{v}_1-\mathbf{v}_2)\sin 2\theta\frac{\partial}{\partial\mathbf{r}}\sigma_{3c}(\mathbf{p}),$$

$$\frac{d}{dt}\sigma_{cu}(\mathbf{p})=ig_0\sigma_{3u}(\mathbf{p})+i\delta_{cu}(\mathbf{v})[\sigma_{cc}(\mathbf{p})-\sigma_{uu}(\mathbf{p})]$$

$$+i\Delta_{cu}(\mathbf{v})\sigma_{cu}(\mathbf{p})+\sin 2\theta\int\Phi(\hat{\mathbf{n}})[\gamma_1\sigma_{33}$$

$$\times(\mathbf{p}-\hbar\mathbf{k}_1+\hat{\mathbf{n}}\hbar k)-\gamma_2\sigma_{33}(\mathbf{p}-\hbar\mathbf{k}_2+\hat{\mathbf{n}}\hbar k)]$$

$$\times d\hat{\mathbf{n}}+\frac{1}{2}(\mathbf{v}_1+\mathbf{v}_2)\frac{\partial}{\partial\mathbf{r}}\sigma_{cu}(\mathbf{p})$$

$$+\frac{1}{4}(\mathbf{v}_1-\mathbf{v}_2)\sin 2\theta\frac{\partial}{\partial\mathbf{r}}[\sigma_{cc}(\mathbf{p})+\sigma_{uu}(\mathbf{p})], \quad (29)$$

where $\mathbf{v}_1=\hbar\mathbf{k}_1/M$ and $\mathbf{v}_2=\hbar\mathbf{k}_2/M$ are the recoil velocities, $|\mathbf{v}_1|\simeq|\mathbf{v}_2|\simeq v_r$, v_r is the total recoil velocity. The effective detunings in Eqs. (29) are defined as

$$\delta_c(\mathbf{v})=(\delta_1-\mathbf{k}_1\cdot\mathbf{v})\sin^2\theta+(\delta_2-\mathbf{k}_2\cdot\mathbf{v})\cos^2\theta+\omega_r,$$

$$\delta_u(\mathbf{v})=(\delta_1-\mathbf{k}_1\cdot\mathbf{v})\cos^2\theta+(\delta_2-\mathbf{k}_2\cdot\mathbf{v})\sin^2\theta+\omega_r,$$

$$\delta_{cu}(\mathbf{v})=\frac{1}{2}[\delta_1-\delta_2+(\mathbf{k}_2-\mathbf{k}_1)\cdot\mathbf{v}]\sin 2\theta=\delta(\mathbf{v})\sin 2\theta,$$

$$\Delta_{cu}(\mathbf{v})=[\delta_1-\delta_2+(\mathbf{k}_2-\mathbf{k}_1)\cdot\mathbf{v}]\cos 2\theta=2\delta(\mathbf{v})\cos 2\theta. \quad (30)$$

Equations (29) are related to the leading function w, which is defined, as usual, by a trace, but over the superpositional states:

$$w(\mathbf{r},\mathbf{p},t)=\sigma_{cc}(\mathbf{r},\mathbf{p},t)+\sigma_{uu}(\mathbf{r},\mathbf{p},t)+\sigma_{33}(\mathbf{r},\mathbf{p},t). \quad (31)$$

This function, which it is natural to call a Wigner quasifunction, satisfies the equation of motion

$$\frac{d}{dt}w=2\int\Phi(\hat{\mathbf{n}})[\gamma_1\sigma_{33}(\mathbf{p}-\hbar\mathbf{k}_1+\hat{\mathbf{n}}\hbar k)+\gamma_2\sigma_{33}$$

$$\times(\mathbf{p}-\hbar\mathbf{k}_2+\hat{\mathbf{n}}\hbar k)]d\hat{\mathbf{n}}-2(\gamma_1+\gamma_2)\sigma_{33}(\mathbf{p})$$

$$+(\mathbf{v}_1\sin^2\theta+\mathbf{v}_2\cos^2\theta)\frac{\partial}{\partial\mathbf{r}}\sigma_{cc}(\mathbf{p})$$

$$+(\mathbf{v}_1\cos^2\theta+\mathbf{v}_2\sin^2\theta)\frac{\partial}{\partial\mathbf{r}}\sigma_{uu}(\mathbf{p})$$

$$+\frac{1}{2}(\mathbf{v}_1-\mathbf{v}_2)\sin 2\theta\frac{\partial}{\partial\mathbf{r}}[\sigma_{cu}(\mathbf{p})+\sigma_{uc}(\mathbf{p})]. \quad (32)$$

Physically, the Wigner quasifunction w is different from the Wigner atomic function since it describes an atomic wave packet whose states are perturbed by the light field (1). At the same time, the normalization condition for the Wigner quasifunction is identical to the normalization condition (25) for the atomic Wigner function:

$$\int w(\mathbf{r},\mathbf{p},t)d^3r\,d^3p=1. \quad (33)$$

The atomic Wigner function expressed in terms of the density matrix elements in the representation of superpositional states is determined by the equation

$$W(\mathbf{r},\mathbf{p},t)=\sin^2\theta[\sigma_{cc}(\mathbf{p}+\hbar\mathbf{k}_1)+\sigma_{uu}(\mathbf{p}+\hbar\mathbf{k}_2)]$$

$$+\cos^2\theta[\sigma_{cc}(\mathbf{p}+\hbar\mathbf{k}_2)+\sigma_{uu}(\mathbf{p}+\hbar\mathbf{k}_1)]$$

$$+\frac{1}{2}\sin 2\theta[\sigma_{cu}(\mathbf{p}+\hbar\mathbf{k}_1)+\sigma_{uc}(\mathbf{p}+\hbar\mathbf{k}_1)$$

$$-\sigma_{cu}(\mathbf{p}+\hbar\mathbf{k}_2)-\sigma_{uc}(\mathbf{p}+\hbar\mathbf{k}_2)]+\sigma_{33}(\mathbf{p}). \quad (34)$$

Equations (29) together with Eq. (32) are a natural set of equations describing effects of optically induced atomic coherence in a three-level Λ-atom. It is clear that, at zero effective detuning $\delta_{cu}(\mathbf{v})=0$ [i.e., when condition (13) is satisfied] and in the case of an infinitely broad atomic wave packet, the third equation in set (29) describes an atomic state which is decoupled from the light field and other atomic states. The decoupling of atomic state $|u\rangle$ from other states means that the whole atomic population is pumped to this optically decoupled state when the detuning is $\delta_{cu}(\mathbf{v})=0$. Thus, the set of equations (29) explicitly describes optical pumping of a velocity-selective CPT state.

Formally, Eqs. (29) describe an effectively three-level atom with upper state $|3\rangle$ and two ground states, $|c\rangle$ and $|u\rangle$ (Fig. 1b). Given the CPT condition (13), Eqs. (29) can be

treated as three coupled groups of equations, namely, equations describing the evolution of an effectively two-level atom with the states $|3\rangle$ and $|c\rangle$, equations describing the optical pumping of state $|u\rangle$ decoupled from the optical field, and equations describing the evolution of coherence between the states of the effectively two-level atom and the decoupled state $|u\rangle$.

4. COUNTER-PROPAGATING LASER WAVES

Integral-differential equations (29) and (32) can be analyzed in many practically important cases. Below we derive as a specific example analytical estimates for the form of the momentum distribution of Λ-atoms at times when a two-peak structure arises due to the CPT effect. For definiteness we consider the case of interaction of spatially broad atomic wave packets with counter-propagating electromagnetic waves, i.e., the case $\mathbf{k}_1=-\mathbf{k}_2\simeq\mathbf{k}=k\mathbf{e}_z$, which has been studied in many experiments. For simplicity we assume equal detunings, $\delta_1=\delta_2=\delta$, equal relaxation rates, $\gamma_1=\gamma_2=\gamma=\Gamma/2$, and the equal Rabi frequencies, $g_1=g_2=g=g_0/\sqrt{2}$.

In the case of a spatially broad atomic wave packet, space derivatives $\partial\sigma_{\alpha\beta}/\partial\mathbf{r}$ in Eqs. (29) and (32) can be neglected. Furthermore, we note that the recoil due to spontaneous photon emission can only spread the momentum distribution. Hence, when the peak widths are comparable to or larger than the recoil momentum, we can replace functions $\sigma_{33}(\mathbf{p}\pm\hat{\mathbf{n}}\hbar k)$ in Eqs. (29) with their approximate expressions obtained by expanding in powers of the recoil momentum. With these simplifications, equations (29) reduce to a set of approximate difference equations

$$\frac{d}{dt}\sigma_{33}=ig_0(\sigma_{c3}-\sigma_{3c})-2\Gamma\sigma_{33},$$

$$\frac{d}{dt}\sigma_{cc}=ig_0(\sigma_{3c}-\sigma_{c3})-ikv(\sigma_{cu}-\sigma_{uc})+\frac{1}{2}\Gamma[\sigma_{33}(\mathbf{p}-\hbar\mathbf{k})+\sigma_{33}(\mathbf{p}+\hbar\mathbf{k})]+\ldots,$$

$$\frac{d}{dt}\sigma_{uu}=ikv(\sigma_{cu}-\sigma_{uc})+\frac{1}{2}\Gamma[\sigma_{33}(\mathbf{p}-\hbar\mathbf{k})+\sigma_{33}(\mathbf{p}+\hbar\mathbf{k})]+\ldots,$$

$$\frac{d}{dt}\sigma_{3c}=ig_0(\sigma_{cc}-\sigma_{33})+(i\delta_r-\Gamma)\sigma_{3c}-ikv\sigma_{3u},$$

$$\frac{d}{dt}\sigma_{3u}=ig_0\sigma_{cu}+(i\delta_r-\Gamma)\sigma_{3u}-ikv\sigma_{3c},$$

$$\frac{d}{dt}\sigma_{cu}=ig_0\sigma_{3u}-ikv(\sigma_{cc}-\sigma_{uu})+\frac{1}{2}\Gamma[\sigma_{33}(\mathbf{p}-\hbar\mathbf{k})-\sigma_{33}(\mathbf{p}+\hbar\mathbf{k})]+\ldots, \tag{35a}$$

and Eq. (32) to the approximate equation

$$\frac{\partial}{\partial t}w=\Gamma[\sigma_{33}(\mathbf{p}-\hbar\mathbf{k})+\sigma_{33}(\mathbf{p}+\hbar\mathbf{k})]-2\Gamma\sigma_{33}+\frac{1}{2}\hbar^2k^2\Gamma\sum\alpha_{ii}\frac{\partial^2}{\partial p_i^2}[\sigma_{33}(\mathbf{p}-\hbar\mathbf{k})+\sigma_{33}(\mathbf{p}+\hbar\mathbf{k})], \tag{35b}$$

where $p=Mv_z$, v_z is the velocity projection on the z-axis, $\Gamma=2\gamma$, and $\delta_r=\delta+\omega_r$. The coefficients α_{ii} $(i=x,y,z)$ are determined by the angular distribution of the rate of spontaneous photon emission:

$$\alpha_{ii}=\int n_i^2\Phi(\hat{n})d\hat{n}. \tag{36}$$

For the practically important case of circularly polarized photons (the $\sigma^+-\sigma^-$ configuration)

$$\Phi(\hat{\mathbf{n}})=\frac{3}{16\pi}(1+n_z^2),\quad \alpha_{xx}=\alpha_{yy}=\frac{3}{10},\quad \alpha_{zz}=\frac{2}{5}.$$

Note that the arguments in Eqs. (35) are only those momenta which are shifted by the recoil momentum.

These simplifications in the equations for the density matrix do not radically change the momentum distribution, since they have no effect on the superpositional state representation. In particular, the simplifications do not affect the splitting of the wave packet in the momentum space determined by Eq. (34).

Further analysis of the equation for the density matrix can be performed for interaction times longer than characteristic evolution times of internal atomic states, but not so long that the scale of deformation of the momentum distribution is smaller than the recoil momentum.

If we neglect the recoil momentum in Eqs. (35a) at these interaction times and calculate steady-state solutions for those density matrix elements that decay with the rate of spontaneous decay, Γ, and then substitute the steady-state solutions in the rest of equations (35a), we find that Eqs. (35a) contain in an implicit form another decay constant of order g^2/Γ. This second constant determines the decay time of coherence between superpositional states. Given the presence of comparable relaxation times in the set of equations, one can neglect in the lowest approximation changes in the momenta due to recoil and calculate steady-state density matrix elements at times $t\gg\Gamma^{-1}$, $(\Gamma/g)^2\Gamma^{-1}$. The latter can be expressed through the Wigner quasifunction (quasiprobability density) $w(\mathbf{p})$:

$$\sigma_{33}(\mathbf{p})=n_3(p)w(\mathbf{p}),\quad \sigma_{cc}(\mathbf{p})=n_c(p)w(\mathbf{p}),$$

$$\sigma_{uu}(\mathbf{p})=n_u(p)w(\mathbf{p}),\quad \sigma_{cu}(\mathbf{p})=\sigma_{uc}^*(\mathbf{p})=\nu(p)w(\mathbf{p}), \tag{37}$$

with the dimensionless functions

$$n_3(p)=G\left(\frac{kv}{\Gamma}\right)^2d(p),$$

$$n_c(p)=\frac{1}{2}\left(\frac{kv}{\Gamma}\right)^2\left[1+G+\left(\frac{kv}{\Gamma}\right)^2+\left(\frac{\delta_r}{\Gamma}\right)^2\right]d(p),$$

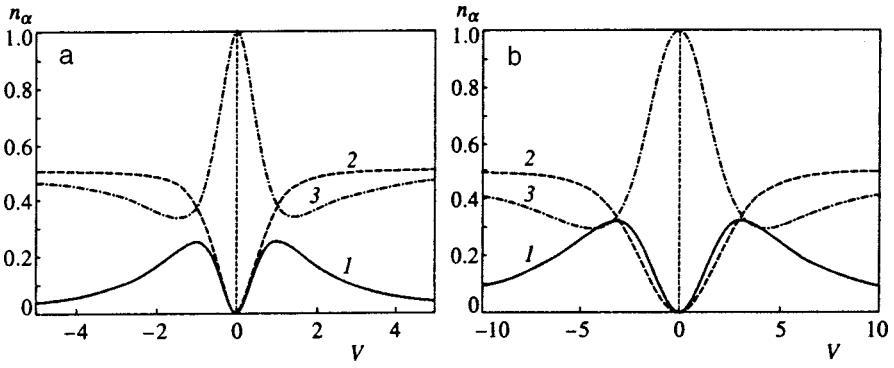


FIG. 2. Relative populations n_α of states (curve *1*) $\alpha=3$, (*2*) $\alpha=c$, and (*3*) $\alpha=u$ as functions of the dimensionless velocity $V=kv/\Gamma$ at detuning $\delta=0$, saturation parameter (a) $G=1$, (b) $G=10$, and $\omega_r/\Gamma=0.05$.

$$n_u(p)=1-\frac{1}{2}\left(\frac{kv}{\Gamma}\right)^2\left[1+3G+\left(\frac{kv}{\Gamma}\right)^2+\left(\frac{\delta_r}{\Gamma}\right)^2\right]d(p),$$

$$\nu(p)=\frac{1}{2}\left(\frac{kv}{\Gamma}\right)\left[\frac{\delta_r}{\Gamma}\left(G-2\left(\frac{kv}{\Gamma}\right)^2+iG\right)\right]d(p) \tag{38}$$

and the resonance factor d and saturation parameter G:

$$d(p)=\left\{G^2+\left(\frac{kv}{\Gamma}\right)^2\left[1+G+\left(\frac{kv}{\Gamma}\right)^2+\left(\frac{\delta_r}{\Gamma}\right)^2\right]\right\}^{-1},$$

$$G=\frac{1}{2}\left(\frac{g_0}{\Gamma}\right)^2=\left(\frac{g}{\Gamma}\right)^2. \tag{39}$$

In the case of a spatially broad atomic wave packet, the functions defined by Eqs. (38) determine relative populations $n_\alpha(p)$ of states $\alpha=3,c,u$ and coherence $\nu(p)$ between the superpositional states (Fig. 2).

It can be seen now from Eq. (34) that in the lowest approximation, the approximate solution (37) defines a two-peak momentum distribution concentrated near the momenta $\pm\hbar k$. In the most interesting region of small momenta $|\Delta p|\leqslant M\Gamma/k$, when $n_3\simeq n_c\simeq\nu\simeq 0$, $n_u\simeq 1$, the shape of the two-peak distribution, in accordance with Eq. (34), is defined by the formula

$$\begin{aligned}W(\mathbf{p},t)=&\frac{1}{2}[\sigma_{uu}(\mathbf{p}+\hbar\mathbf{k})+\sigma_{uu}(\mathbf{p}-\hbar\mathbf{k})+\sigma_{cc}(\mathbf{p}+\hbar\mathbf{k})\\&+\sigma_{cc}(\mathbf{p}-\hbar\mathbf{k})+\sigma_{cu}(\mathbf{p}+\hbar\mathbf{k})+\sigma_{uc}(\mathbf{p}+\hbar\mathbf{k})\\&-\sigma_{cu}(\mathbf{p}-\hbar\mathbf{k})-\sigma_{uc}(\mathbf{p}-\hbar\mathbf{k})]+\sigma_{33}(\mathbf{p})\\\simeq&\frac{1}{2}[w(\mathbf{p}+\hbar\mathbf{k})+w(\mathbf{p}-\hbar\mathbf{k})].\end{aligned} \tag{40}$$

It follows from Eq. (40) that an effective estimate of the time evolution of the Wigner function W can be obtained using the equation for the time evolution of the leading function w. By substituting the solution (37) in the lowest approximation in Eq. (35b) and neglecting the momentum shift in the quasiprobability density, one can obtain an approximate equation for $w=w(\mathbf{p},t)$:

$$\frac{\partial}{\partial t}w=Qw+\sum\frac{\partial^2}{\partial p_i^2}(D_{ii}w). \tag{41}$$

The functions Q and D_{ii} in Eq. (41) defined by equations

$$Q=\Gamma[n_3(p-\hbar k)+n_3(p+\hbar k)]-2n_3(p), \tag{42}$$

$$D_{ii}=\hbar^2k^2\Gamma\alpha_{ii}n_3(p), \tag{43}$$

determine the effective pumping rate and quasidiffusion tensor.

Equation (41) has a direct physical interpretation. Depending on the sign of Q, the first term on the right-hand side of Eq. (41) describes either the optical pumping of the decoupled state $|u\rangle$ or depopulation of the decoupled state. The term with the derivatives describes the diffusive broadening of peaks on the momentum scale owing to spontaneous decays. Since the optical pumping term includes input terms $n_3(v\pm v_r)$ at shifted velocities, according to Eq. (42), the optical pumping is effective only in a narrow range of small velocities. The latter statement means that the population of the $|u\rangle$ state is gradually transferred to a narrow peak around the zero velocity, and this process, in accordance with Eq. (40) determines the two-peak shape of the velocity distribution.

The dependence of the effective pumping rate on the atomic velocity, which controls the process of optical pumping of the CPT state, is illustrated by Fig. 3. In accordance with Eq. (42), the $|u\rangle$ state is optically populated in a narrow range of velocities determined by the recoil velocity. Outside this narrow range where $Q<0$ holds, the population of the decoupled state $|u\rangle$ drops to zero with time. The effective pumping rate as a function of atomic velocity at different light wave intensities is plotted in Fig. 4. At low wave intensities the pumping rate is a steep function of the atomic velocity because of a narrow dip in the population n_3 of the upper level as a function of velocity. At higher intensities of

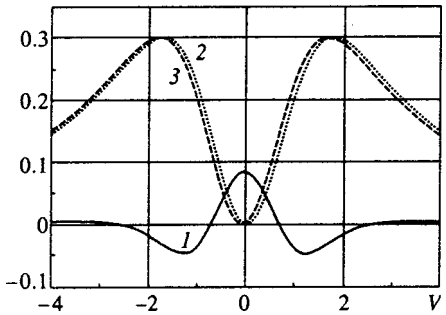


FIG. 3. Normalized effective optical pumping rate $S=50Q/\Gamma$ (curve *1*) and relative populations of the re- and blue-shifted upper state: (*2*) $n_3(v+v_r)$ and (*3*) $n_3(v-v_r)$, as functions of the dimensionless velocity $V=kv/\Gamma$ at detuning $\delta=0$, saturation parameter $G=3$, and $\omega_r/\Gamma=0.05$.

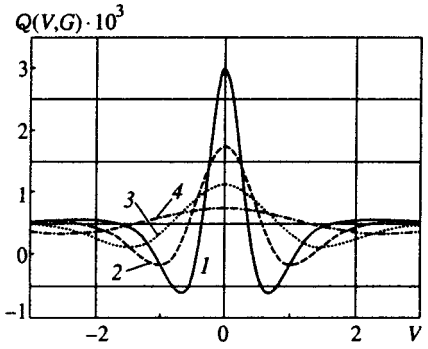


FIG. 4. Normalized effective optical pumping rate $Q(v,G)=Q/2\Gamma$ as a function of dimensionless velocity $V=kv/\Gamma$ at $\delta=0$, $\omega_r/\Gamma=0.05$, and (curve *1*) $G=1$; (*2*) 2; (*3*) 4; (*4*) 10.

light waves, both the dip in the upper state population and the peak of the effective pumping rate versus atomic velocity are broadened.

As follows from Eq. (41), the typical time in which two peaks are created is inversely proportional to both the light wave intensity and recoil energy squared:

$$\tau=\frac{1}{Q(0)}\simeq\frac{1}{8G}\left(\frac{\Gamma}{\omega_r}\right)^2\Gamma^{-1}.$$

For example, in the case of ^{4}He atoms driven by two circularly polarized laser waves at the $2^3S_1-2^3P_1$ transition ($\lambda=1.08\ \mu$m, $\omega_r=2.7\times10^5\ \mathrm{s}^{-1}$, $2\Gamma=1.6\times10^6\ \mathrm{s}^{-1}$) at intermediate values of the saturation parameter $G=(g/\Gamma)^2\simeq1$, the characteristic time is 8×10^{-6} s. This estimate is in good agreement with experimental data.[8]

V. CONCLUSIONS

The theoretical approach to optical pumping of decoupled CPT states in three-level Λ-atoms described in this paper has demonstrated that it is natural to apply the density matrix in the superpositional state representation to the optical formation of velocity-selective CPT states. Our investigation has clearly shown that the superpositional state representation allows one to describe directly velocity-selective CPT states and to derive a kinetic equation for the quasiprobability density, which determines the formation and evolution of a two-peak velocity distribution. The kinetic description of the evolution of the velocity (momentum) distribution shows that formation of a velocity-selective CPT state can be interpreted in terms of optical pumping under conditions of weak quasidiffusion. The kinetic approach to the problem developed on this basis allows one to derive parameters of the process leading to formation of a velocity-selective CPT state from a single basic parameter, namely the effective optical pumping rate.

The research was supported by the International Science Foundation (Grant NL1000), the International Science Foundation and the Government of Russia (Grant NL1300), and the Russian Fund for Fundamental Research (Grant 95-02-05350).

*)E-mail: minogin@isan.troitsk.ru

[1] E. Arimondo and G. Orriols, Lett. Nuovo Cimento **17**, 33 (1976).
[2] H. M. Gray, R. M. Whitley, and C. R. Stroud, Opt. Lett. **3**, 218 (1978).
[3] F. T. Hioe, J. Opt. Soc. Am. B **4**, 1327 (1987).
[4] F. T. Hioe and C. E. Carrol, Phys. Rev. A **37**, 3000 (1988).
[5] V. S. Smirnov, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **96**, 1613 (1989) [Sov. Phys. JETP **69**, 913 (1989)].
[6] M. Kaivola, P. Thorsen, and O. Poulsen, Phys. Rev. A **32**, 207 (1985).
[7] M. S. Shuhriar and P. R. Hemmer, Phys. Rev. Lett. **65**, 1865 (1990).
[8] A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, Phys. Rev. Lett. **61**, 826 (1988).
[9] A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, J. Opt. Soc. Am. B **6**, 2112 (1989).
[10] V. A. Alekseev and D. D. Krylova, Lazernaya Fizika **2**, 781 (1992).
[11] F. Bardou, J. P. Bouchaud, O. Emile, A. Aspect, and C. Cohen-Tannoudji, Phys. Rev. Lett. **72**, 203 (1994).
[12] J. Lawall, S. Culin, B. Saubamea, N. Bigelow, M. Leduc, and C. Cohen-Tannoudji, Phys. Rev. Lett. **75**, 4194 (1995).
[13] O. Kocharovskaya, S.-Y. Zhu, M. O. Scully, P. Mandel, and Y. Y. Radeonychev, Phys. Rev. A **49**, 4928 (1994).
[14] V. G. Minogin and V. S. Letokhov, *Laser Field Pressure on Atoms* [in Russian], Nauka, Moscow (1987).

Translation provided by the Russian Editorial office.

Quantum cryptography using photon ''frequency'' states (example of a possible realization)

S. N. Molotkov*)

Institute of Solid-State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

(Submitted 27 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 526–537 (August 1998)

A quantum cryptosystem is proposed using single-photon states with different frequency spectra as information carriers. A possible experimental implementation of the cryptosystem is discussed. © *1998 American Institute of Physics.* [S1063-7761(98)01208-6]

1. INTRODUCTION

The idea of quantum cryptography was first proposed in Ref. 1, initially inaccessible. In its current form the protocol for propagation of a key (a secret random sequence of zeros and ones) was proposed in Ref. 2. A new qualitative jump in the understanding of secrecy in quantum cryptography arose after Ref. 3, in which a protocol was proposed for exchange using nonorthogonal states, and Ref. 4 where a protocol was described using the Einstein–Podolsky–Rosen effect.

Various versions of cryptosystems using nonorthogonal states and the Einstein–Podolsky–Rosen effect were proposed subsequently.[5–22] Experimental prototypes of quantum cryptosystems have been implemented using nonorthogonal polarization states of photons for coding[5,11,12] and also using the principle of phase coding based on a fiber-optic Mach–Zehnder interferometer with time separation[8–10,19] (an improved version of the system from Ref. 10 was implemented in Ref. 19). The longest communication channel achieved under laboratory conditions is 30 km (Ref. 10). The operation of a prototype quantum cryptosystem has been demonstrated under natural conditions using a 23 km long optical cable below Lake Geneva[18] and also between various buildings at the Los Alamos National Laboratory.[13]

Most of these systems use the interference principle which roughly involves ''splitting'' a photon at the transmitting end of the line and ''collimating'' it at the receiving end. Here we propose a quantum cryptosystem using different photon frequency states which utilizes the ''internal interference'' of the different photon frequency components. Such a system may well prove more stable in operation than direct interference systems, although this can only be confirmed by means of an experimental implementation.

The secrecy of the key in quantum cryptography is based on two facts: 1) the impossibility of coding (cloning) a previously unknown quantum state[23] and 2) the impossibility of extracting information on quantum states without perturbing them if they belong to a nonorthogonal basis.[3] Formally, any pair of nonorthogonal states corresponding to logic 0 and 1 can be used as information carriers. The detection procedure (quantum-mechanical measurement) at the receiving end should be set up so that any attempts at intervention within the communication channel, i.e., changes in states, can be identified from the results of the measurements. If the pair of nonorthogonal states $|\psi_0\rangle$ and $|\psi_1\rangle$ are used as carriers, the formal measurements are given by the projectors[3]

$$\overline{E}_0 = 1 - |\psi_0\rangle\langle\psi_0|, \quad \overline{E}_1 = 1 - |\psi_1\rangle\langle\psi_1|,$$

whose action reduces to the projection of states orthogonal to the vectors $|\psi_0\rangle$ and $|\psi_1\rangle$, respectively, on the subspaces.[3] The result of the action of the projectors is treated as a statement and the probability of the measurement results is given by the expressions

$$\Pr = \mathrm{Tr}\{\hat{\rho}_0\overline{E}_0\} = \mathrm{Tr}\{\hat{\rho}_1\overline{E}_1\} \equiv 0,$$

$$\Pr = \mathrm{Tr}\{\hat{\rho}_0\overline{E}_1\} = \mathrm{Tr}\{\hat{\rho}_1\overline{E}_0\} = 1 - |\langle\psi_0|\psi_1\rangle|^2 \neq 0. \qquad (1)$$

Measurements using $\overline{E}_0$ and $\overline{E}_1$ in an ideal communication channel (without noise) can detect any attempts at eavesdropping, i.e., changes in states. The first nonzero outcome of a control measurement definitely indicates the presence of an eavesdropper. When it is known that a signal, say $|\psi_0\rangle$, was sent, measurements were made using $\overline{E}_0$, and a nonzero result was obtained, the nonzero result is considered to be a statement that the state $\overline{E}_0$ has a nonzero component in the corresponding orthogonal complement of the Hilbert space (i.e., that the state $|\psi_0\rangle$ was changed).

Photon states are used as information carriers for real fiber-optic communication channels. Formally any pair of nonorthogonal photon states can be used (not necessarily even single-photon states). However, it is unclear how the measurement procedure corresponding to the projectors $\overline{E}_{0,1}$ for a given pair of states can be implemented experimentally. It would be easiest to use a pair of nonorthogonal polarizations, but an optical fiber does not maintain polarization (see details given in Refs. 11 and 12). The prototypes of quantum cryptosystems mentioned use the phase coding principle based on interferometers with time separation.[8–10] After the interferometer has been operating for a few minutes, the system requires additional alignment.[10]

2. CRYPTOGRAPHY USING PHOTON FREQUENCY STATES

A quantum cryptographic system using the Einstein–Podolsky–Rosen effect for a biphoton field was proposed in

an earlier study[20] and this idea is developed here. Subsequently, the use of frequency states which do not use interference over large distances is suggested.

Let us first analyze a formal system and then discuss an experimental realization. Three single-photon states are used as carriers: two information states corresponding to logic zero and one and one control state. The information states are mutually orthogonal. The control state is pairwise nonorthogonal to the information states. The use of only two information states is inadequate for secrecy because they can be reliably distinguished as a result of their orthogonality.

The information states consist of pure stationary states with the density matrices

$$\hat{\rho}_0=|e_0\rangle\langle e_0|, \quad \hat{\rho}_1=|e_1\rangle\langle e_1|, \quad \langle e_1|e_0\rangle=0, \tag{2}$$

where $|e_0\rangle$ and $|e_1\rangle$ are certain basis states assigned to the energies ω_0 and ω_1, respectively. The control state is nonstationary and contains both basis components $|e_0\rangle$ and $|e_1\rangle$:

$$|\psi_c(t_0)\rangle=\mathrm{e}^{-i\omega_0 t_0}f_0|e_0\rangle+\mathrm{e}^{-i\omega_1 t_0}f_1|e_1\rangle,$$

$$\hat{\rho}_c(t_0)=|\psi_c(t_0)\rangle\langle\psi_c(t_0)|, \tag{3}$$

and the normalization condition

$$|f_0|^2+|f_1|^2=1.$$

The time t_0 describes the beginning of the time measurement, the state preparation time (see below). The density matrix at times $t>t_0$ is obtained by substituting into the argument $\hat{\rho}_c(t)=|\psi_c(t-t_0)\rangle\langle\psi_c(t-t_0)|$. The introduction of two orthogonal information states reduces the number of ''dummy'' outcomes because of their distinguishability if no eavesdropping is detected in the exchange process.

This scheme uses two types of measurements. The measurements of the frequency spectrum are described by an orthogonal partition of unity in space relative to the states $|e_0\rangle$, $|e_1\rangle$:

$$E_0+E_1=I, \quad E_0=|e_0\rangle\langle e_0|, \quad E_1=|e_1\rangle\langle e_1|, \tag{4}$$

where I is the unit operator. The second family of measurements involves measuring the time and is given by a nonorthogonal partition of unity (see Ref. 24, for example), which in our case has the form

$$\int_0^T E(dt)=I, \quad T=\frac{2\pi}{|\omega_1-\omega_0|},$$

$$E(dt)=(\mathrm{e}^{-i\omega_0 t}|e_0\rangle+\mathrm{e}^{-i\omega_1 t}|e_1\rangle)(\langle e_0|\mathrm{e}^{i\omega_0 t}+\langle e_1|\mathrm{e}^{i\omega_1 t})\frac{dt}{T}. \tag{5}$$

In accordance with the general philosophy of quantum-mechanical measurements, the measurements are made at a particular time.[24–26] The probability of the outcome of the measurements using the projectors E_0 and E_1 does not depend on time and is given by

$$\mathrm{Pr}=\mathrm{Tr}\{\hat{\rho}_0E_0\}=1, \quad \mathrm{Pr}=\mathrm{Tr}\{\hat{\rho}_1E_1\}=1,$$

$$\mathrm{Pr}=\mathrm{Tr}\{\hat{\rho}_{0,1}E_{1,0}\}\equiv 0,$$

$$\mathrm{Pr}=\mathrm{Tr}\{\hat{\rho}_c(t)E_0\}=|f_0|^2, \quad \mathrm{Pr}=\mathrm{Tr}\{\hat{\rho}_c(t)E_1\}=|f_1|^2. \tag{6}$$

Measurements of the time give the probability distribution of the outcomes in the range $(t,t+dt)$:

$$\mathrm{Pr}(dt)=\mathrm{Tr}\{\hat{\rho}_{0,1}E(dt)\}=1\cdot\frac{dt}{T}, \tag{7}$$

$$\mathrm{Pr}(dt)=\mathrm{Tr}\{\hat{\rho}_c(t)E(dt)\}=|f_0\exp[-i\omega_0(t-t_0)]+f_1\exp[-i\omega_1(t-t_0)]|^2\left(\frac{dt}{T}\right)$$

$$=\{1+2\mathrm{Re}[f_0f_1^*\exp[-i(\omega_0-\omega_1)\times(t-t_0)]]\}\left(\frac{dt}{T}\right). \tag{8}$$

For the control state the probability is an oscillating function with the period $T=2\pi/|\omega_1-\omega_0|$. This set of measurements can completely reconstruct information on the states—no other density matrices can reproduce the statistics of the measurements so that any attempts at eavesdropping can be detected (for further details see Ref. 22).

The key generation protocol is as follows. We assume that all the parameters of the states are known to everybody, including any potential eavesdropper. User A (henceforth ''Alice'') randomly sends into the communication channel states $\hat{\rho}_c$, $\hat{\rho}_0$, or $\hat{\rho}_1$. User B (henceforth ''Bob'') randomly and independently of Alice selects measurement type E_0, E_1, or $E(dt)$. After making a series of measurements, Alice transmits through the open channel (accessible to all including an eavesdropper,''Eve'') the numbers of some measurements when $\hat{\rho}_0$ and $\hat{\rho}_1$ were sent and all the numbers when control state $\hat{\rho}_c$ was sent. Bob sorts the measurements into three groups according to when $\hat{\rho}_c$, $\hat{\rho}_0$, or $\hat{\rho}_1$ were transmitted. In each of these three groups, three subgroups are identified according to measurement procedures E_0, E_1, or $E(dt)$. For instance, for those messages when Alice transmitted state $\hat{\rho}_c$, the relative fraction of the measurement outcomes when the projectors E_0 and E_1 were used should be $|f_0|^2/|f_1|^2$ regardless of the measurement time. For the $E(dt)$ measurements the probability of the measurement results at various times should converge to the probability distribution (8). The convergence of the distribution function for a finite sample should by checked by using some statistical criterion such as the Kolmogorov criterion[27] (see also Ref. 22). The convergence is checked similarly for the measurements when states $\hat{\rho}_0$ or $\hat{\rho}_1$ were sent. For example, for state $\hat{\rho}_0$ the measurements using E_0 should give the same outcome in all attempts, which does not depend on the measurement time. For the E_1 measurements in all attempts the outcome should be zero regardless of the measurement time. For the $E(dt)$ measurements the probability of the outcome is only determined by the duration of the time interval dt and does not depend on the time t.

The secrecy of the protocol is guaranteed by the nonorthogonality of the information states to the control state and by the fact that a set of measurements is information-complete so that any attempts at eavesdropping, i.e., changes

in states, can be detected. In other words, no other density matrices can reproduce the statistics of the measurements at the receiving end (for further details see Ref. 22).

After having established that no eavesdropping is taking place, Alice transmits the numbers of those measurements when the control state was sent. All the idle measurements when the detector was not actuated are discarded. Then, for the remaining numbers Bob only transmits the numbers of those measurements in which he used E_0 or E_1 but does not communicate which measurement, E_0 or E_1, was used in each specific attempt (this information is now known only to Alice and Bob). These remaining measurements give the secret key (an identical random sequence of zeros and ones for Alice and Bob).

We shall illustrate why an eavesdropper will inevitably introduce errors. In order to obtain information on the key, Eve must distinguish states $\hat{\rho}_0$ and $\hat{\rho}_1$. To do this, she must make measurements with a narrow-band detector (measurements of E_0 or E_1). If there were no control state $\hat{\rho}_c$ containing both spectral components with frequencies ω_0 and ω_1, as a result of the mutual orthogonality of the information states, it would be possible to determine uniquely which state is present in the line. However, their nonorthogonality to the control state will inevitably lead to errors, since there will always be measurements with an undetermined result. For instance, if $\hat{\rho}_c$ is present in the line and Eve measured E_0 and obtained a nonzero result, it is impossible to uniquely determine which state, $\hat{\rho}_c$ or $\hat{\rho}_0$, gave this result. Resending $\hat{\rho}_0$ instead of the true state $\hat{\rho}_c$ leads to a change in Bob's measurement statistics. It is also impossible to discern in one measurement that both spectral components with energies ω_0 and ω_1 are present simultaneously in a state because of the orthogonality of the components, since this requires measurements by means of E_0E_1. This projector can be considered as confirmation that the property E_0 (ω_0 present) and E_1 (ω_1 present) are found simultaneously. However because of the orthogonality ($E_0 \cap E_1 = \varnothing$) the action of E_0E_1 on any density matrix has a result identically equal to zero. Also, no unique information can be obtained on the simultaneous presence of spectral components using more general (non-von Neumann) measurements, which is guaranteed by the theorem in Ref. 3.

3. POSSIBLE IMPLEMENTATION OF A CRYPTOSYSTEM

We shall now discuss a possible experimental implementation in which the carriers are three single-photon states of the form

$$|1_{\omega_0}\rangle = a^+_{\epsilon,\omega_0}|0\rangle, \quad |1_{\omega_1}\rangle = a^+_{\epsilon,\omega_1}|0\rangle,$$
$$|1_c\rangle = f_0 e^{-i\omega_0 t_0} a^+_{\epsilon,\omega_0}|0\rangle + f_1 e^{-i\omega_1 t_0} a^+_{\epsilon,\omega_1}|0\rangle \tag{9}$$

with the corresponding density matrices

$$\hat{\rho}_{0,1} = |1_{\omega_{0,1}}\rangle\langle 1_{\omega_{0,1}}|, \quad \hat{\rho}_c = |1_c\rangle\langle 1_c|,$$

where a^+_{ϵ,ω_i} is the creation operator of a Fock monochromatic state with the frequency ω_i ($i=0,1$) and polarization ϵ, and $|0\rangle$ is the vacuum state. Quite clearly, a strictly monochromatic state is an idealization. However, there are no fundamental constraints on the formation of states arbitrarily close to monochromatic.

The measurement procedures described above may be implemented by using a fast (fairly wide-band) photodetector operated in a waiting regime, and two narrow-band filters at frequencies ω_0 and ω_1. From standard photodetection theory,[28] the detection probability is proportional to the first-order correlation function of the field

$$\Gamma^{(1)}(t) = \mathrm{Tr}\{\hat{\rho}_i \hat{E}^{(-)}(x,t)\hat{E}^{(+)}(x,t)\}, \tag{10}$$

where

$$\hat{E}^{(+)}(x,t) = i\sum_{\omega_n} \sqrt{\frac{\hbar\omega_n}{2V}}\, a_{\epsilon_n,\omega_n} \exp(-i\omega_n t + ik_n x),$$

$$\hat{E}^{(-)}(x,t) = -i\sum_{\omega_n} \sqrt{\frac{\hbar\omega_n}{2V}}\, a^+_{\epsilon_n,\omega_n} \exp(i\omega_n t - ik_n x),$$

and V is the normalization volume. At this stage it is more convenient to use a formal normalization of the states in a finite volume (see below). We can even use unnormalized states. With this definition the probabilities of the measurement outcomes will also be unnormalized, but since only the relative probability is important for the different measurements, this lack of normalization is unimportant.

Measurements of the correlation function of the field (the instantaneous intensity) $\Gamma^{(1)}(t)$ are a realization of the $E_{0,1}$ and $E(dt)$ measurements described above in the sense that the statistics of the outcomes gives the same information on the states as the statistics of the $E_{0,1}$ and $E(dt)$ measurements. A combination of measurements using a fast photodetector and measurements using two narrow-band filters and the same photodetector can provide information on the amplitude $|f_{0,1}|$ and relative phase of the components f_0 and f_1, which exhausts the information on the states (see also Ref. 22).

The probability p of a photon being recorded in the time interval $(t, t+dt)$ by an ideal photodetector is proportional to the field intensity $I(t) \propto \Gamma^{(1)}(t)$ (Ref. 28):

$$p(t)dt \propto I(t)dt = \Gamma^{(1)}(t)dt. \tag{11}$$

If the photodetector activation time is $\tau_{\mathrm{det}} \ll 1/|\omega_1 - \omega_0|$, this photodetector implements $E(dt)$ measurements in the sense indicated above. It can be seen from Eq. (10) that for state (9) the recording probability with allowance for Eqs. (9)–(11) has the form

$$p(t)dt \propto I(t)dt = \Gamma^{(1)}(t)dt = \left| \sqrt{\omega_0} f_0 \exp\left[-i\omega_0(t-t_0) + \frac{ik_0 L}{c}\right] + \sqrt{\omega_1} f_1 \exp\left[-i\omega_1(t-t_0) + \frac{ik_1 L}{c}\right]\right|^2 \frac{dt}{2V}, \tag{12}$$

where $k_{0,1}$ are the wave vectors corresponding to the frequencies $\omega_{0,1}$ and L is the length of the communication channel (we assume that the measurement is made at a distance L from the transmitting end).

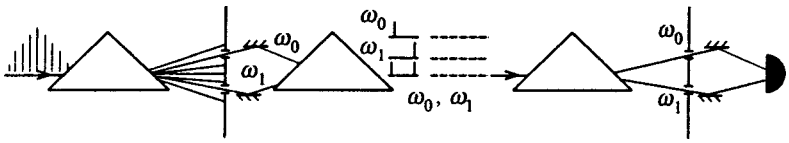


FIG. 1. Schematic of cryptosystem. The signal from a single-photon source is directed to a ''prism'' beyond which is a screen that transmits the signal either with the frequency ω_0 (logic zero, upper diaphragm open) or ω_1 (logic one, lower diaphragm open), or both frequencies (control signal, both diaphragms open). At the receiving end the measurement procedure is arranged similarly. Upper diaphragm open, measurement of E_0; lower diaphragm open, measurement of E_1; and both open, measurement of $E(dt)$. The diaphragms are open or shut during an entire specific message.

Measurements of the amplitude of the spectral components $f_{0,1}$ are made using a pair of narrow-band filters, which cut out the frequencies $\omega_{0,1}$ prior to photodetection, and the same photodetector. The recording probability, in accordance with Eqs. (9)–(11), does not depend on time:

$$p_c(t)dt \propto \Gamma^{(1)}(t)dt = \begin{cases} \dfrac{\hbar\omega_0}{V}|f_0|^2 dt, & \text{measured} \quad E_0, \\ \dfrac{\hbar\omega_1}{V}|f_1|^2 dt, & \text{measured} \quad E_1, \end{cases}$$

$$p_{0,1}(t)dt \propto \Gamma^{(1)}(t)dt = \begin{cases} \dfrac{\hbar\omega_0}{V}\times 1\, dt, & E_0 \quad \text{measured for } \hat{\rho}_0, \\ 0 & E_1 \quad \text{measured for } \hat{\rho}_0, \\ \dfrac{\hbar\omega_1}{V}\times 1\, dt, & E_1 \quad \text{measured for } \hat{\rho}_1, \\ 0 & E_0 \quad \text{measured for } \hat{\rho}_1. \end{cases} \tag{13}$$

A schematic of a cryptosystem is shown in Fig. 1. Before entering the line, the signal from a single-photon source is expanded into a spectrum from which are cut either one of the frequencies (ω_0 or ω_1), or both spectral components with the frequencies ω_0 and ω_1. The $E(dt)$ measurements are made using a fast photodetector operating in a waiting mode. In this mode the occurrence of an event (its recording) will take place at a random time, not chosen by the experimentalist. This differs from the $E(dt)$ measurements made at a time preselected by the experimentalist in the range $(t,t+dt)$; the probability of recording at this time is described by the density $p_c(t)$. In this case, the $E(dt)$ measurement cannot be understood as a measurement by the fast photodetector which has an input diaphragm in front of it which is opened in the interval $(t,t+dt)$. This procedure also corresponds to some measurement, but not to an $E(dt)$ measurement.

The integrated recording probability at time T is given by

$$P(T) = \int^T dt\, p_i(t), \quad i = c, 0, 1,$$

from which the probability density in the time interval $(t,t+dt)$ can be obtained by differentiating with respect to the upper limit.

The $E(dt)$ measurement essentially contains information on the ''interference'' of the different spectral component within a single quantum state (information on the relative phase of the components with the frequencies ω_0 and ω_1). Thus, it is fundamentally important that in different messages the state $\hat{\rho}_c$ is prepared so that the relative phase of the spectral components is the same. Otherwise, the time-oscillating (interference) component with the frequency $\omega_1 - \omega_0$ in the probability $p_c(t)$ will not be reproduced at the same times in different attempts. The problem of preparing a single-photon state where the relative phase of the components is fixed can be solved as follows. Let us assume that a two-level system exists with a spin-nondegenerate electron spectrum (for example, a quantum dot with Coulomb interaction; see details given in Ref. 29). Resonant illumination with a square-wave pulse can transfer the system to the excited (quasi-steady) state. We shall assume that the square-wave pulse is much shorter than the radiative recombination time ($\tau_\pi \ll \tau_R$). The square-wave pulse can be made substantially shorter than τ_R with some margin for not violating the condition of resonant illumination.[29] This implies that the time of excitation t_0 is determined to within $\sim \tau_\pi \ll \tau_R$. After the square-wave pulse has been switched off, the free evolution of the system comprising an electron in the excited state plus the electromagnetic field in the vacuum state leads to recombination of the electron and the appearance of a single-photon packet with the characteristic spectral width $\Delta\omega \approx 1/\tau_R$. The single-photon packet is defined as[30,31]

$$|1_f\rangle = \sum_k f_k a^+_{\omega_k}|0\rangle, \quad \sum_k |f_k|^2 = 1. \tag{14}$$

The average number of photons in the packet is

$$n = \langle 1_f | a^+_{\omega_k} a_{\omega_k} | 1_f \rangle = 1, \tag{15}$$

and physically this implies that recording by an ideal wide-band photodetector (which captures all the spectral components) leads to actuation with a probability of unity. Recording by an ideal narrow-band detector at the frequency ω_n leads to a probability $|f_n|^2 < 1$ of actuation.

Since each system at time t_0 starts from the same state, in different messages the single-photon packets are the same (the phase of all the spectral components determined by the factors $\exp(-i\omega_i t_0)$ is the same in different messages). Cutting out two narrow spectral components from the spectrum conserves their relative phase. In fact, the cutting of the spectral components is formally described as the action of the projector[1)]

$$E_0 + E_1 = (|1_{\omega_0}\rangle\langle 1_{\omega_0}| + |1_{\omega_1}\rangle\langle 1_{\omega_1}|),$$

after which the density matrix of the single-photon wave packet is transferred to a new state (see Refs. 25, 26, and 32)

$$\hat{\rho}_{\text{in}}(t)=\left\{\sum_k \exp[-i\omega_k(t-t_0)]f_k|1_{\omega_k}\rangle\right\}$$
$$\times\left\{\sum_{k'}\langle 1_{\omega'_k}|f^*_{k'}\exp[i\omega'_k(t-t_0)]\right\}$$
$$\to\frac{1}{\text{Tr}\{\hat{\rho}_{\text{in}}(t)(E_0+E_1)\}}(E_0+E_1)$$
$$\times\left\{\sum_k \exp[-i\omega_k(t-t_0)]f_k|1_{\omega_k}\rangle\right\}$$
$$\times\left\{\sum_{k'}\langle 1_{\omega'_k}|f^*_{k'}\exp[i\omega'_k(t-t_0)]\right\}(E_0+E_1)$$
$$\to\frac{1}{|f_0|^2+|f_1|^2}\{\exp[-i\omega_0(t-t_0)]f_0|1_{\omega_0}\rangle$$
$$+\exp[-i\omega_1(t-t_0)]f_1|1_{\omega_1}\rangle\}$$
$$\times\{\langle 1_{\omega_0}|f^*_0\exp[i\omega_0(t-t_0)]$$
$$+\langle 1_{\omega_1}|f^*_1\exp[i\omega_1(t-t_0)]\}. \quad (16)$$

Physically this implies that if an ideal wide-band photodetector is placed after the filters, in a large number of repeated tests it will only be actuated in the fraction $\text{Tr}\{\hat{\rho}_{\text{in}}(t)(E_0+E_1)\}$ of the total number of cases.

The relative phase of the components with frequencies ω_0 and ω_1 is determined by their phase at the preparation time which is attainable in principle, as described above. Thus, provided that $\tau_\pi \ll \tau_R \ll 1/|\omega_1-\omega_0|$ holds, we can assume that in different messages the temporal interference pattern stays in place. In different messages the interference pattern only ''floats'' by virtue of the inaccuracy of the initial preparation time δt_0 by the amount $\delta t_0 \leqslant \tau_\pi \ll T=2\pi/|\omega_1-\omega_0|$, which is substantially less than the period of the temporal interference pattern.

Assuming that the square-wave pulse duration is $\tau_\pi \sim 10^{-12}$ s (see Ref. 29) and the radiative recombination time is $\tau_R \sim 10^{-10}$ s (in this case the spectral width of the initial single-photon state is $\Delta\omega \sim 10^{10}$ Hz), and spectral components of width $\sigma \approx 10^7$ Hz separated by the distance $\delta\omega=|\omega_1-\omega_0|\sim 10^8$ Hz are cut out (which is still very far from the limits now attainable), the required photodetector response speed is satisfied for $\tau_{\text{det}}\sim 10^{-9}$ s.

The chain of inequalities $\tau_\pi \ll \tau_R \ll \tau_{\text{det}} \ll 1/\delta\omega$ is then satisfied with some margin. The efficiency as a result of cutting out narrow spectral components of width $\sigma \sim 10^7$ Hz from a spectrum of width $\Delta\omega \sim 10^{10}$ Hz is $\sim \sigma/\Delta\omega \sim 10^{-3}$. However, at least the problem of a strictly single-photon source can be solved in principle.

We shall estimate the accuracy in fixing the length of the communication channel. Changes in the length of the fiber-optic line also lead to blurring of the interference pattern as a result of the presence of terms with $k_{0,1}L$ in the exponent in formula (12). The changes in the relative phase of the spectral components as a result of variation of the line length δL should satisfy the condition

$$|k_1-k_0|\,\delta L \approx |\omega_1-\omega_0|\,\delta L/c \ll 2\pi,$$

and permissible variations of the line length should lead to a relative phase shift much less than 2π. This gives the estimate

$$\delta L \ll 2\pi c/\delta\omega \approx 10^2 \text{ cm},$$

which is a fairly mild condition.

The interference pattern may also become blurred as a result of the polarization vector rotating at different speeds in the different frequency components. However, if the position of the cable is fixed, the interference pattern may be precalibrated. In this case, any changes will only be attributed to the different optical paths of the frequency components, i.e., variation of the line length. This condition is clearly noncritical. The frequency dispersion of the dielectric constant of the optical fiber can also lead to smoothing of the amplitude of the oscillations of the interference pattern. The longer the line, the stronger this smoothing. However, estimates[21] show that if the width of the spectral components is $\sigma \approx 10^7$ Hz, the dispersion has an influence at far greater lengths than the attenuation. Attenuation does not influence the secrecy of the system and only reduces its efficiency by increasing the fraction of idle measurements.

The states with infinitely narrow spectral components analyzed above are an idealization and are unsuitable for transmission along a communication channel because of their formally infinite duration. In real experiments we can only prepare states with a finite line width (the preparation of strictly monochromatic photon states would require a formally infinite time). The information states can be single-photon states of the form (14) with Gaussian spectral densities

$$|1_{\omega_{0,1},c}\rangle=\int_0^\infty f_{0,1,c}(\omega)a^+(\omega)|0\rangle,$$

$$[a(\omega),a^+(\omega')]=\delta(\omega-\omega')\hat{I}, \quad (17)$$

$$E^{(+)}(x,t)=\frac{1}{\sqrt{2\pi}}\int_0^\infty \exp\left[-i\omega\left(t-\frac{x}{c}\right)\right]a(\omega)d\omega, \quad (18)$$

$$f_{0,1}(\omega)=\frac{1}{(2\pi\sigma^2)^{1/4}}\exp\left[-\frac{(\omega-\omega_{0,1})^2}{4\sigma^2}\right]\exp(-i\omega t_0) \quad (19)$$

and a control state containing both narrow-band Gaussian components with amplitudes f_0 and f_1:

$$f_c(\omega)=\frac{\text{const}}{(2\pi\sigma^2)^{1/4}}\left\{f_0\exp\left[-\frac{(\omega-\omega_0)^2}{4\sigma^2}\right]\right.$$
$$\left.+f_1\exp\left[-\frac{(\omega-\omega_1)^2}{4\sigma^2}\right]\right\}\exp(-i\omega t_0), \quad (20)$$

where the normalization constant is

$$\text{const}=\left\{|f_0|^2+|f_1|^2+\sqrt{2}\,\text{Re}[f_0 f_1^*]\right.$$
$$\left.\times\exp\left[-\frac{(\omega_0^2+\omega_1^2-\omega_0\omega_1)}{2\sigma^2}\right]\right\}^{-1}. \quad (21)$$

Measurements of narrow spectral components using suitable Gaussian filters yield a weak time dependence of the measurement outcomes unlike the previous analysis of strictly monochromatic states, where the probability of the outcome did not depend on time. The corresponding probability density of the results has the form

$$p(t)dt\propto I(t)dt=2\sqrt{2\pi\sigma^2}\exp[-2\sigma^2(t-t_0-L/c)^2]dt, \quad (22)$$

from which it follows in particular that the probability of recording by a photodetector in the waiting mode only tends to unity if the waiting time T exceeds the reciprocal width of the spectrum ($T\geqslant 1/\sigma$). This factor is consistent with intuitive ideas on the prolonged recording time of a narrow-band state.

The probability density of the measurement outcomes for the control state has the form

$$p_c(t)dt\propto I(t)dt=\text{const}\cdot 2\sqrt{2\pi\sigma^2}\exp[-2\sigma^2$$
$$\times(t-t_0-L/c)^2]\{|f_0|^2+|f_1|^2+2\,\text{Re}(f_0 f_1^*$$
$$\times\exp[-i(\omega_0-\omega_1)(t-t_0-L/c)])\}dt. \quad (23)$$

The interference oscillating component is well-defined under the condition $\sigma\ll|\omega_1-\omega_0|$.

4. CONCLUSIONS

Although various versions of quantum cryptosystems have been proposed, in the author's view there is some indeterminacy associated with the following. The proof of secrecy in quantum cryptography using two nonorthogonal states proposed in Ref. 3 implies that the states are stationary and belong to the same energy. Otherwise, for nonstationary states the projectors $\overline{E_0}$ and $\overline{E_1}$ would differ at different times. The nonorthogonality of the stationary states implies that they correspond to the same energy. Otherwise the stationary states belonging to different energies would automatically be orthogonal. In this sense, the protocol for the stationary states exists, as it were, outside time. Attempts to introduce the time explicitly in the exchange protocol[14,17] still use reasoning from Ref. 3 for stationary states as proof of secrecy (see, for example, Ref. 17). The stationary state are infinite in time. A similar situation arises here. Evidence of secrecy for states with infinitely narrow spectral densities, which are thus infinitely extended in time, is also based on the reasoning given in Ref. 3. For the case where the state space of the system is infinite-dimensional (described by a continuous variable), proof of secrecy of the same degree of rigor as in Ref. 3 is not obtained, as far as we are aware. In this sense, the author takes the view that evidence of secrecy is not obtained for real-time quantum cryptosystems.

In conclusion, the author is grateful to B. A. Volkov, S. S. Nazin, S. T. Pavlov, and I. I. Tartakovskiĭ for fruitful discussions. This work was supported by the Russian Fund for Fundamental Research (Project No. 96-02-19396).

*)E-mail:molotkov@issp.ac.ru

[1)]Strictly speaking, an ideal filter corresponds to a projector on the subspace of states with frequencies ω_0, ω_1 and all occupation numbers, $E_{\omega_0}+E_{\omega_1}=\Sigma_{n=0}^{\infty}(|n_{\omega_0}\rangle\langle n_{\omega_0}|+|n_{\omega_1}\rangle\langle n_{\omega_1}|)$, but this does not alter the results.

[1] S. Wiesner, *Conjugate Coding*, Sigact News. **15**, 78 (1983); original text dated 1970.
[2] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing*, Bangalore, India (IEEE, New York, 1984), p. 175.
[3] C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992); C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).
[4] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[5] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, J. Cryptology **5**, 3 (1992).
[6] A. K. Ekert, J. G. Rarity, P. R. Tapster, and G. M. Palma, Phys. Rev. Lett. **69**, 1293 (1992).
[7] A. Muller, J. Brequet, and N. Gisin, Europhys. Lett. **30**, 809 (1994).
[8] R. J. Hughes, D. M. Alde, P. Dyer, G. G. Luther, G. L. Morgan, and M. Schauer, Contemp. Phys. **36**, 149 (1995).
[9] S. J. D. Phoenix and P. D. Townsend, Contemp. Phys. **36**, 165 (1995).
[10] C. Marand and P. D. Townsend, Opt. Lett. **20**, 1695 (1995).
[11] J. D. Franson and H. Ilves, Appl. Opt. **33**, 2949 (1994).
[12] J. D. Franson and H. Ilves, J. Mod. Opt. **41**, 2391 (1994).
[13] R. J. Hughes, G. G. Luther, G. L. Morgan, C. G. Peterson, and C. Simmons, in *Advances in Cryptography–Proceedings of Crypto'96*, August 1996 (Springer-Verlag, Berlin, 1996).
[14] L. Goldenberg and L. Vaidman, Phys. Rev. Lett. **75**, 1239 (1995).
[15] M. Koashi and N. Imoto, Phys. Rev. Lett. **77**, 2137 (1996).
[16] E. Biham, B. Huttner, and T. Mor, Phys. Rev. A **54**, 2651 (1996).
[17] M. Koashi and N. Imoto, Phys. Rev. Lett. **79**, 2383 (1997).
[18] A. Muller, H. Zbinden, and N. Gisin, Nature (London) (London) **378**, 449 (1995).
[19] A. Muller, T. Herzog, B. Huttner, W. Tittel, H. Zbinden, and N. Gisin, Appl. Phys. Lett. **70**, 793, (1997).
[20] S. N. Molotkov and S. S. Nazin, JETP Lett. **62**, 273 (1995).
[21] S. N. Molotkov and S. S. Nazin, JETP Lett. **63**, 687 (1996); **64**, 866 (1996)]; S. N. Molotkov, JETP Lett. **64**, 702 (1996); **65**, 586 (1997); **66**, 773 (1997).
[22] S. N. Molotkov and S. S. Nazin, JETP Lett. **66**, 779 (1997).
[23] W. K. Wootters and W. H. Zurek, Nature (London) **299**, 802 (1982).
[24] A. S. Kholevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Elsevier, New York, 1982) [Russian original, Nauka, Moscow 1980].
[25] J. Von Neumann, *Mathematical Foundations of Quantum Mechanics*, transl. from the German (Princeton University Press, Princeton, N.J., 1955) [Nauka, Moscow, 1964].
[26] P. Busch, M. Grabowski, and P. J. Lahti, *Operational Quantum Physics*, Springer Lecture Notes in Physics, Vol. 31 (1995).
[27] A. N. Kolmogorov, Giorn. Ist. Attuari **4**, 83 (1933); see also P.-L. Hennequin and A. Tortrat, *Theorie des Probabilites et quelques Applications*, Collection d'ouvrages de mathematiques a l'usage des physiciens, Masson, Paris (1965) [Nauka, Moscow (1974)].
[28] J. Perina, *Coherence of Light* (Van Nostrand Reinhold, New York, 1972) [Mir, Moscow, 1974].
[29] F. V. Krasheninnikov, L. A. Openov, S. N. Molotkov, and S. S. Nazin, Zh. Éksp. Teor. Fiz. **112**, 1257 (1997) [JETP **85**, 682 (1997)].
[30] R. J. Glauber, Phys. Rev. **130**, 2529 (1963).
[31] R. A. Campos, B. E. Salech, and M. Teich, Phys. Rev. A **42**, 4127 (1990).
[32] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976) [Mir, Moscow, 1979].

Translated by R. M. Durham

Theory of spin diffusion in liquid-phase polymer systems

N. F. Fatkullin and G. A. Yatsenko*)

Kazan State University, 420008 Kazan, Russia

R. Kimmich and E. Fischer

Sektion Kernresonanzspektroskopie, Universität Ulm, 89069 Ulm, Germany

(Submitted 24 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 538–554 (August 1998)

A general theory of spin diffusion in condensed media is constructed by the method of Zwanzig–Mori projection operators using the superpositional approximation to decouple the many-particle correlation functions. The spin diffusion coefficient is expressed in the form $D_{sp}=D_{tr}+D_f$, where D_{tr} is the contribution associated with translational displacements of the molecules and D_f is the contribution caused by intermolecular flip-flop processes. The expression for D_{tr} differs from the well-known Kubo–Green formula for the self-diffusion coefficient D_{sd} in that the integrand contains an additional factor $P_f(t)$, which is the probability of the molecular spins not participating in intermolecular flip-flop transitions over the time t. A microscopic expression is obtained for D_f in the form of a time integral of the intermolecular dipole–dipole dynamic correlation functions. For liquid-phase polymer system with fairly high molecular mass the condition $D_{sp}\gg D_{sd}$ is satisfied. © *1998 American Institute of Physics.* [S1063-7761(98)01308-0]

1. INTRODUCTION

The phenomenon of spin diffusion is well-known in NMR spectroscopy of solids (see, for example, Refs. 1–6). It is the basic mechanism for the establishment of a spatially uniform distribution for an arbitrary longitudinal magnetizing field $\mu_z(\mathbf{r},t)$. Magnetic dipole–dipole interactions between spins of different unit cells induce flip-flop transitions which result in spatial transport of the spin polarization, which helps to equalize $\mu_z(\mathbf{r},t)$ in different regions of the solid.

The characteristic size of the spatial nonuniformities of the field $\mu_z(\mathbf{r},t)$ is generally much greater than the characteristic linear dimension a_0 of a unit cell:

$$\left|a_0\frac{\partial}{\partial r}\mu_z(\mathbf{r},t)\right|\ll|\mu_z(\mathbf{r},t)|,$$

and the characteristic times t satisfy the inequality $\tau_f\ll t\ll T_1$, where τ_f is the characteristic time of the intercell flip-flop transitions and T_1 is the spin–lattice relaxation time. If these conditions are satisfied, the evolution of the field $\mu_z(\mathbf{r},t)$ is determined by the classical diffusion equation:

$$\frac{\partial}{\partial t}\mu_z(\mathbf{r},t)=D_{sp}\Delta\mu_z(\mathbf{r},t), \tag{1}$$

where D_{sp} is the spin diffusion coefficient, Δ is the Laplacian, and $\mathbf{r}$ is the spatial position vector.

The order-of-magnitude spin diffusion coefficient is estimated as

$$D_{sp}\simeq\frac{1}{6}\frac{a_0^2}{\tau_f}. \tag{2}$$

In magnetically concentrated solids the time τ_f is approximately 10^{-4} s and $a_0\approx 3$ Å, which gives $D_{sp}\simeq 10^{-16}$ m^2/s.

In liquid-phase systems a spatially nonuniform field $\mu_z(\mathbf{r},t)$ can also be created by stimulated spin echo with a pulsed magnetic field gradient.[7–9] Intermolecular dipole–dipole interactions are capable of inducing intermolecular flip-flop transitions. The characteristic times of these transitions $\tau_f=10^{-2}-1$ s are much greater because the mobility of the molecules in the liquid is higher than that in solids. Nevertheless, on the experimentally attainable time intervals $\tau_f\ll t\ll T_1$ the evolution of the field $\mu_z(\mathbf{r},t)$ satisfies Eq. (1).

However, in this case, the coefficient D_{sp} differs negligibly from the self-diffusion coefficient of spin-carrying molecules: $D_{sd}\simeq D_{sp}$. This is attributed to the high translational mobility of liquid molecules which move in space over distances of the order of 10^2-10^4 Å or more within times of the order τ_f, with the lower limit even being typical of extremely viscous macromolecular melts. Thus, at first glance, it can be shown that for liquids flip-flop transitions cannot have any significant influence on the rate at which the spatially nonuniform field $\mu_z(\mathbf{r},t)$ becomes uniform. Thus, there is no reason to distinguish between the self-diffusion coefficient of the molecules D_{sd} and the spin diffusion coefficient D_{sp}.

Arguments were put forward earlier to suggest that for macromolecular solutions and melts with fairly high molecular masses, the situation is not so simple.[10–12] This is because the maximum relaxation time of macromolecules depends strongly on the molecular mass: $\tau_{max}\propto\tau_s N^{3.4}$, where $\tau_s=10^{-11}-10^{-9}$ s is the segmental relaxation time and N is the number of Kuhn segments per macromolecule. At the same time, the molecular mass dependence of the time τ_f

should be similar to that of the spin–spin relaxation time T_2, which initially decreases with increasing N and then reaches a constant value,[13] i.e., becomes independent of the molecular mass of the macromolecules. The time τ_f should satisfy the conditions $T_2 < \tau_f < T_1$, so that the estimate $\tau_f = 10^{-2}$–10^{-1} s is reasonable. Thus, there is a certain critical molecular mass N^* for which

$$\tau_f = \tau_s N^{*3.4}. \tag{3}$$

We then obtain $N^* = (\tau_f/\tau_s)^{1/3.4} \approx 10^2 - 10^3$.

It is found[12] that the relation between the coefficients of spin diffusion and self-diffusion depends strongly on the relation between the critical molecular mass N^* and the molecular mass of the polymer melt N. For $N \ll N^*$ both these coefficients are almost equal, whereas for $N \gg N^*$ the inequality $D_{sp} \gg D_{sd}$ is satisfied. We shall demonstrate this by citing one of the results of Ref. 12:

$$D_{sp} = \frac{1}{6\tau_f}[\langle r^2(\tau_f)\rangle + a_0^2], \tag{4}$$

where $\langle r^2(\tau_f)\rangle$ is the mean-square displacement of the macromolecular segments over the time τ_f and a_0^2 is the additional mean-square displacement of the spin polarization caused by intermolecular flip-flop processes.

In solids we find $\langle r^2(\tau_f)\rangle^{1/2} \approx 0.01$ Å and $a_0 \approx 3$ Å, i.e., $a_0^2 \gg \langle r^2(\tau_f)\rangle$, so that Eq. (4) is in fact equivalent to expression (2). In liquids with molecular masses $N \ll N^*$ for times $t \gg \tau_{max}$ the mean-square molecular displacements satisfy the normal diffusion conditions (Einstein relation):

$$\langle r^2(t)\rangle = 6D_s t. \tag{5}$$

Since the value of a_0 does not exceed the linear dimensions of the molecule (macromolecule), it follows from Eqs. (5) and (4) that

$$D_{sp} = \frac{1}{6\tau_f}(6D_s\tau_f + a_0^2) \simeq D_s. \tag{6}$$

However, in the limit $N \gg N^*$ relation (5) ceases to be valid, since for times of order τ_f the macromolecular motion resembles anomalous diffusion. The mean-square displacement of the segments may be estimated as

$$\langle r^2(t)\rangle \approx b^2 N^{-\alpha}(t/\tau_s)^{\beta}, \tag{7}$$

where the exponents α and β are determined by details of the micromolecular dynamics and b is the length of a Kuhn segment.

For the Rouse model, for example, we have $\alpha = 1/2$, $\beta = 1/2$ (Ref. 14), for the reptation model we have $\alpha = 0$, $\beta = 1/4$ in region II and $\alpha = 1/2$, $\beta = 1/2$ in region III (Ref. 14), and for the renormalized Rouse model, $\alpha = 0$, $\beta = 2/5$ (Refs. 15 and 16), and so on.

An alternative description of anomalous diffusion to relation (7) is the concept of an effective, generally time-dependent, self-diffusion coefficient, defined as

$$D_{sd}^*(t) = \frac{1}{6}\frac{\langle r^2(t)\rangle}{t} \approx \frac{1}{6}\frac{b^2}{N^{\alpha}}\frac{1}{\tau_s^{\beta} t^{1-\beta}}, \quad t \ll \tau_{max}. \tag{8}$$

The effective self-diffusion coefficient decreases with time for $t \ll \tau_{max}$ and $D_{sd}^*(t) \gg D_{sd}$. Using Eqs. (7) and (4), we obtain the inequalities

$$D_{sp} \geqslant D_{sd}^*(\tau_f) \gg D_{sd}, \tag{9}$$

which are valid even when the contribution containing a_0^2 is neglected in relation (6). The substantial difference between the coefficients of self-diffusion and spin diffusion in liquid polymer systems with $N \gg N^*$ has a simple physical explanation.

Spin diffusion takes place as a result of spatial displacements of elementary spin polarization quanta. In the limit $N \ll N^*$, $\tau_{max} \ll \tau_f$, randomization, i.e., transfer of the spatial trajectories of the spin polarization quanta to the asymptotic diffusion curve, takes place within times τ_{max}, i.e., much earlier than the flip-flop transitions and thus $D_{sd} \simeq D_{sp}$. In the limit $N \gg N^*$, $\tau_f \ll \tau_{max}$ the trajectories of the spin polarization quanta become randomized at times τ_f, whereas the trajectories of the spin-carrying polymer segments become randomized within times of the order τ_{max}, i.e., much later, so that $D_{sp} \gg D_{sd}$.

In order to avoid confusion over terminology, we shall allow ourselves some additional explanations. Strictly speaking, spin diffusion should be understood as the recovery of spatially uniform magnetization in a system removed from a state of thermodynamic equilibrium regardless of its phase state: solid, liquid, or gas. This is the effect considered in fairly general studies (see Ref. 17, for example). In this case, the coefficient of spin diffusion is the kinetic coefficient D_{sp} in Eq. (1). This naturally gives rise to the question: ''in which cases are the coefficients of self-diffusion (molecular diffusion) and spin diffusion similar and when do they differ substantially?''

In the vast majority of studies dealing with specific systems, the term ''spin diffusion'' is associated with solids and the coefficient of spin diffusion is identified with the second term in Eq. (4), i.e., with expression (2). Although this is not the case, strictly speaking, the smallness of the first term in relation (4) does not yield any actual errors. In liquids, as far as we are aware, it was assumed until the publication of Refs. 10–12 that the diffusion coefficient measured by the method of stimulated spin echo with a pulsed magnetic field gradient is always the same as the self-diffusion coefficient of spin-carrying molecules. For the reasons put forward above serious differences may arise here. We shall call these spin diffusion effects or effects of intermolecular flip-flop processes.

To the best of our knowledge, the first observation of anomalously high diffusion coefficients measured experimentally by the method of stimulated spin echo with a pulsed magnetic field gradient was reported in the experimental studies of Res. 18,19 Although the authors did not attribute these experimental results to spin diffusion effects, they did stimulate the theoretical predictions published in Refs. 10–12. The results published in Ref. 20 also indicate that macromolecules may have anomalously high measurable coefficients of self-diffusion, although this observation lay outside the scope of their investigation. Recently, the existence of spin diffusion effects in macromolecular melts

was demonstrated experimentally by comparing the diffusion coefficients measured by the method of stimulated spin echo with a pulsed magnetic field gradient in melts of deuterated and nondeuterated polyethyleneoxide molecules with molecular mass $M=450\,000$ (Ref. 21). The observation of spin diffusion effects in polystyrene solutions was also reported in Ref. 22. In the light of this new experimental situation further theoretical studies of this problem are required.

The result (4) was in fact obtained by us earlier using the density matrix method for the spin subsystem of a sample and treating the operator of the magnetic dipole–dipole interaction energy $\hat{V}$ as a perturbation. Actual calculations were only made with allowance for effects quadratic in $\hat{V}$ and accurately describing the behavior of the spin system only on time scales $t \ll \tau_f$. The result (4) is only meaningful for times $t \gg \tau_f$. Thus, strictly speaking, it was not derived but was inferred by extrapolating the results of the lowest order of perturbation theory. It is therefore natural to attempt to obtain the result (4) without recourse to perturbation theory and provide a more reliable basis for the microscopic calculations of the parameters τ_f and a_0.

The present paper reports a systematic discussion of this topic.

2. DERIVATION OF A GENERAL RELATION FOR THE SPIN DIFFUSION COEFFICIENT

This section contains the derivation of the transport equations describing spin diffusion processes in a certain sample placed in a static, generally speaking, spatially nonuniform external magnetic field. A convenient means of achieving this aim is the method of Zwanzig–Mori projection operators (see Refs. 17, 23, and 24) by means of which one can derive the so-called generalized Langevin equation.

As far as we are aware, the derivation of the generalized Langevin equation describing the kinetics of the longitudinal components of the magnetizing field has not been examined in the literature with the generality required for our purposes, although a particular case was studied in Ref. 17 where the magnetic dipole–dipole interactions were neglected and only the translational displacements of the spins were taken into account. Thus, for greater cohesiveness we shall describe the necessary underlying assumptions of the method of Zwanzig–Mori projection operators as well as details of the derivation of the transport equation of interest to us.

2.1. Generalized Langevin equation

We shall analyze two arbitrary physical quantities described by the quantum-mechanical operators $\hat{A}$ and $\hat{B}$. Between these quantities we can determine the scalar product in the Kubo sense:

$$\langle \hat{A}|\hat{B}\rangle \equiv \beta^{-1}\int_0^\beta d\lambda \,\mathrm{Tr}\,(\hat{A}^*\hat{B}(i\hbar\lambda)\hat{\rho}_{\mathrm{eq}}), \tag{10}$$

where $\hat{A}^*$ is an operator which is the Hermitian conjugate of $\hat{A}$,

$$\hat{B}(\tau)\equiv \exp\left\{ i\frac{\hat{H}\tau}{\hbar}\right\}\hat{B}\exp\left\{ -i\frac{\hat{H}\tau}{\hbar}\right\}$$

is the operator $\hat{B}$ in the Heisenberg representation, $\beta = 1/kT$ is the reciprocal temperature, and $\hat{\rho}_{\mathrm{eq}}$ is the equilibrium density matrix of the entire system.

The set of all linear operators treated as a linear space with the metric defined by relation (10) forms the so-called Liouville space L. Subsequently we shall use as the operators $\hat{A}$ and $\hat{B}$ some specially selected linear combination of spin operators whose time evolution is significant for times of order $\tau \geqslant 10^{-2}$ s. The characteristic time $\hbar\lambda$ which appears in the definition of the scalar product (10) is of the of order $\hbar/kT \simeq 10^{-13}$ s $\ll \tau_f$. This allows us to simplify relation (10):

$$\langle \hat{A}|\hat{B}\rangle = \mathrm{Tr}\,(\hat{A}^*\hat{B}\hat{\rho}_{\mathrm{eq}}) = \langle \hat{A}^*\hat{B}\rangle_{\mathrm{eq}}, \tag{11}$$

where $\langle \hat{A}^*\hat{B}\rangle_{\mathrm{eq}}$ denotes the equilibrium autocorrelation function of $\hat{A}^*$ and $\hat{B}$.

Let us assume that we are interested in the kinetics of a certain set of physical quantities $A_1, A_2, \ldots, A_n$. The set of all linear combinations of these quantities forms the linear subspace $L^n\{A_i\} \in L$. The existence of a metric in the entire Liouville space L allows us to determine the projection operator on $L^n\{A_i\}$:

$$\hat{P} \equiv \sum_{k,m} |A_k\rangle \langle A|A\rangle^{-1}_{km}\langle A_m|, \tag{12}$$

where $\langle A|A\rangle^{-1}_{km}$ is a matrix element of the matrix which is the inverse of the static correlation matrix $\langle A|A\rangle \equiv ||\langle A_k|A_m\rangle||$, $|A_m\rangle$ is the ket vector in the Liouville space L corresponding to the operator $\hat{A}_m$, and $\langle A_k|$ is a bra vector in the Liouville space L corresponding to the operator $\hat{A}_k$.

Mori transformations can be used to formally obtain an accurate system of equations for the operators $\hat{A}_n(t)$ in the Heisenberg representation:

$$\frac{d}{dt}\hat{A}_n(t) = \sum_k i\omega_{nk}\hat{A}_k(t) - \sum_k \int_0^t d\tau K_{nk}(\tau)\hat{A}_k(t-\tau) + F^Q_n(t), \tag{13}$$

where

$$\omega_{nk} = \sum_m \frac{1}{\langle A|A\rangle_{km}}\langle A_m|\hat{L}|A_n\rangle \tag{14}$$

is the frequency matrix, $\hat{L} \equiv (1/\hbar)[\hat{H}, \ldots]$ is the Liouville operator of the system,

$$F^Q_n(t) \equiv \exp\{i\hat{Q}\hat{L}t\}i\hat{Q}\hat{L}|A_n\rangle \tag{15}$$

is a generalized stochastic Langevin force associated with the physical quantity $\hat{A}_n$, $\hat{Q} = 1 - \hat{P}$ is the projection operator on the subspace orthogonal to $L^n\{\hat{A}_i\}$, and

$$K_{nk} \equiv \sum_m \frac{1}{\langle A|A\rangle_{km}}\langle F^Q_m(0)|F^Q_n(\tau)\rangle \tag{16}$$

is the memory matrix.

2.2. Generalized Langevin equation for the longitudinal magnetizing field

We shall define the microscopic magnetizing field as

$$\hat{\mu}_z(\mathbf{r}) \equiv \sum_{i=1}^{N} \mu_0 \hat{I}_i^z \delta(\mathbf{r}-\mathbf{r}_i), \tag{17}$$

where μ_0 is the magnetic moment of the spin, $\hat{I}_i^z$ is the z component of the operator of the ith spin, $\mathbf{r}$ is the ''field'' position vector, and $\mathbf{r}_i$ is the position vector of the ith spin.

Using a Fourier representation for the Dirac δ function in relation (17), we obtain the following expression for the magnetizing field:

$$\hat{\mu}_z(\mathbf{r}) = \int \frac{d\mathbf{k}}{(2\pi)^3} \exp\{-i\mathbf{k}\cdot\mathbf{r}\} \left(\sum_{i=1}^{N} \exp\{i\mathbf{k}\cdot\mathbf{r}_i\} \mu_0 \hat{I}_i^z \right). \tag{18}$$

We define the collective (''hydrodynamic'') modes of this field characterized by the wave vector $\mathbf{k}$ as follows:

$$\hat{\mu}_z(\mathbf{k}) = \mu_0 \sum_{i=1}^{N} \hat{I}_i^z \exp\{i\mathbf{k}\cdot\mathbf{r}_i\}. \tag{19}$$

A simple relationship exists between the field $\hat{\mu}_z(\mathbf{r})$ and the collective modes $\hat{\mu}^z(\mathbf{k})$:

$$\hat{\mu}_z(\mathbf{k}) = \int \frac{d\mathbf{k}}{(2\pi)^3} \exp\{-i\mathbf{k}\cdot\mathbf{r}_i\} \hat{\mu}^z(\mathbf{k}). \tag{20}$$

In the following calculations, the collective variables $\hat{\mu}^z(\mathbf{k})$ will play the role of the quantities $\hat{A}_n$, and the wave vector $\mathbf{k}$ plays the role of the index n numbering the quantities $\hat{\mu}^z(\mathbf{k})$. The transport equations for these are obtained from the equations (12) by replacing summation over the discrete index n by integration over the continuous index $\mathbf{k}$.

We shall first calculate the scalar products of two modes with the wave vectors $\mathbf{k}_1$ and $\mathbf{k}_2$:

$$\langle \hat{\mu}^z(\mathbf{k}_1) | \hat{\mu}^z(\mathbf{k}_2) \rangle = \mu_0^2 \sum_{i,j} \langle \hat{I}_i^z \hat{I}_j^z \rangle_{\mathrm{eq}} \langle \exp\{i\mathbf{k}_2\cdot\mathbf{r}_j - i\mathbf{k}_1\cdot\mathbf{r}_i\} \rangle_{\mathrm{eq}}. \tag{21}$$

In the high-temperature approximation in terms of the spin variables, no spin–spin correlations exist:

$$\langle \hat{\mu}_i^z \hat{\mu}_j^z \rangle_{\mathrm{eq}} = \frac{\mu_0^2 I(I+1)}{3} \delta_{ij}. \tag{22}$$

Substituting this relation into expression (21), we obtain

$$\langle \hat{\mu}^z(\mathbf{k}_1) | \hat{\mu}^z(\mathbf{k}_2) \rangle = \frac{\mu_0^2 I(I+1)}{3} \rho_s (2\pi)^3 \delta(\mathbf{k}_1 - \mathbf{k}_2), \tag{23}$$

where ρ_s is the spin density in this particular sample.

Thus, the normal modes $\hat{\mu}^z(\mathbf{k})$ are mutually orthogonal. The right-hand side of Eq. (23) may be considered to be a matrix element of a matrix with the continuous indices $\mathbf{k}_1$ and $\mathbf{k}_2$. Its inverse matrix $\langle \hat{\mu}^z \otimes \hat{\mu}^z \rangle^{-1}_{\mathbf{k}_1\mathbf{k}_2}$ is given by

$$\int d\mathbf{k}_3 \langle \hat{\mu}^z \otimes \hat{\mu}^z \rangle^{-1}_{\mathbf{k}_1\mathbf{k}_3} \langle \hat{\mu}(\mathbf{k}_3) | \hat{\mu}^z(\mathbf{k}_2) \rangle \equiv \delta(\mathbf{k}_1 - \mathbf{k}_2). \tag{24}$$

This definition and formula (23) give

$$\langle \hat{\mu}^z \otimes \hat{\mu}^z \rangle^{-1}_{\mathbf{k}_1\mathbf{k}_2} = \frac{3}{\mu_0^2 I(I+1) \rho_s (2\pi)^3} \delta(\mathbf{k}_1 - \mathbf{k}_2). \tag{25}$$

Then, using this equality and the general definition (12), we obtain the following expression for the projection operator on the half-space extended to the hydrodynamic variables $\hat{\mu}^z(\mathbf{k})$:

$$\hat{P} = \frac{3}{(2\pi)^3} \frac{1}{\mu_0^2 I(I+1)\rho_s} \int d\mathbf{k} |\hat{\mu}^z(\mathbf{k})\rangle\langle\hat{\mu}^z(\mathbf{k})|. \tag{26}$$

The following calculations require some refinement of the structure of the system Hamiltonian:

$$\hat{H} = \hat{H}_S + \hat{H}_L + \hat{V}, \tag{27}$$

where $\hat{H}_S$ is the ''spin Hamiltonian'' describing the interaction of the spins with the external magnetic fields, $\hat{H}_L$ is the lattice Hamiltonian which describes the spatial displacements of the molecules and atoms in the system, and $\hat{V}$ is the ''spin–lattice interaction'' operator, which in our case is the same as the Hamiltonian of the dipole–dipole interactions.

In the cases of interest to us, the external magnetic field may be represented as the sum of two terms oriented along the z axis: a constant term and a spatially nonuniform gradient term. Accordingly, the spin Hamiltonian contains two contributions:

$$\hat{H}_S = \sum_n \hbar \omega_0 \hat{I}_n^z - \sum_n \hbar \gamma g z_n \hat{I}_n^z, \tag{28}$$

where ω_0 is the Larmor precession frequency, g is the magnetic field gradient, γ is the spin gyromagnetic ratio, and z_n is the spatial z coordinate of the nth spin.

In the Hamiltonian $\hat{V}$ it is most interesting for our purposes to express the so-called Van Vleck B component which describes flip-flop processes in the spin subsystem:

$$\hat{V}_f = \sum_{k<m} b_{km} (\hat{I}_k^+ \hat{I}_m^- + \hat{I}_k^- \hat{I}_m^+), \tag{29}$$

where

$$b_{km} = -\frac{1}{4} \frac{\gamma^4 \hbar^2}{r_{km}^3} (1 - 3\cos^2\theta_{km}),$$

$\mathbf{r}_{km}$ is the position vector connecting spins numbers k and m, and θ_{km} is the polar angle between $\mathbf{r}_{km}$ and the z axis.

We shall now derive the generalized Langevin equation for the hydrodynamic modes $\hat{\mu}^z(\mathbf{k})$ of a magnetizing field having the following structure (compare with the general relation (13)):

$$\frac{d}{dt}\hat{\mu}^z(\mathbf{k},t) = \int d\mathbf{k}' \Omega(\mathbf{k}|\mathbf{k}') \hat{\mu}^z(\mathbf{k},t) - \int_0^t d\tau \times \int d\mathbf{k} M(\mathbf{k}|\mathbf{k}';\tau) \hat{\mu}^z(\mathbf{k}, t-\tau) + f^{zQ}(\mathbf{k},t), \tag{30}$$

where $\Omega(\mathbf{k}|\mathbf{k}')$ is the frequency matrix which is the analog of ω_{nk} in formula (13), $M(\mathbf{k}|\mathbf{k}';\tau)$ is the memory matrix which is the analog of K_{nk} in formula (13), and $f^{zQ}(\mathbf{k},t)$ is the stochastic Langevin force associated with the mode $\hat{\mu}^z(\mathbf{k})$, which is the analog of $F_n^Q(t)$.

Using formulas (14), (19), and (28), we can easily confirm that in our case

$$\Omega(\mathbf{k}|\mathbf{k}')=0. \tag{31}$$

In order to calculate $f^{zQ}(\mathbf{k},t)$, we need to use formula (15) and the analogies described above:

$$\hat{f}^{zQ}(\mathbf{k},t)\equiv\exp\{i\hat{Q}\hat{L}t\}i\hat{Q}\hat{L}|\hat{\mu}^z(\mathbf{k})\rangle. \tag{32}$$

We shall first calculate this force for the initial time:

$$\hat{f}^{zQ}(\mathbf{k})=i\hat{Q}\hat{L}\hat{\mu}^z(\mathbf{k})=\frac{i}{\hbar}\hat{Q}[\hat{H};\hat{\mu}^z(\mathbf{k})]$$

$$=\frac{i}{\hbar}[\hat{H}_L;\hat{\mu}^z(\mathbf{k})]+\frac{i}{\hbar}[\hat{V};\hat{\mu}^z(\mathbf{k})]. \tag{33}$$

It then becomes clear that initially the operator $f^{zQ}(\mathbf{k},t)$ is the same as the operator of the moment of force acting on the hydrodynamic mode of the magnetizing field $\hat{\mu}^z(\mathbf{k})$, taken in the Schrödinger representation.

The first term in relation (33) is associated with the spatial displacement of the spins for which we introduce the special notation:

$$\hat{f}^z_{\rm tr}(\mathbf{k})\equiv\frac{i}{\hbar}[\hat{H}_L;\hat{\mu}^z(\mathbf{k})]. \tag{34}$$

The second term describes the transport of the spin polarization induced by flip-flop processes, and we denote this by

$$\hat{f}^z_f(\mathbf{k})\equiv\frac{i}{\hbar}[\hat{V}_f;\hat{\mu}^z(\mathbf{k})]. \tag{35}$$

We shall calculate the translational contribution to the moment of force:

$$\hat{f}^z_{\rm tr}(\mathbf{k})\equiv\frac{i}{\hbar}\left[\sum_n\frac{\hat{\mathbf{p}}_n^2}{2m};\sum_k\exp\{i\mathbf{k}\cdot\mathbf{r}_k(0)\}\hat{I}^z_k\right]$$

$$\simeq\sum_m\hat{\mathbf{V}}_m(0)\cdot\mathbf{k}\exp\{i\mathbf{k}\cdot\mathbf{r}_m(0)\}\hat{I}^z_m. \tag{36}$$

Note that in going over to the last relation, we neglected quantum-mechanical effects associated with the noncommutating behavior of the operators $\hat{\mathbf{p}}_m$ and $\mathbf{r}_m$. This is permissible since the typical temperatures for liquid-phase polymer systems are $T\simeq 300$ K.

Using formulas (35) and (29), we transform the expression for $\hat{f}^z_f(\mathbf{k})$ to give

$$\hat{f}^z_f(\mathbf{k})=\frac{i}{\hbar}\sum_{k\neq m}b_{km}(0)[1-\exp\{i\mathbf{k}\cdot\mathbf{r}_{km}(0)\}]$$

$$\times\hat{I}^+_k\hat{I}^-_m\exp\{i\mathbf{k}\cdot\mathbf{r}_k(0)\}. \tag{37}$$

We shall analyze the long-wavelength approximation $\mathbf{k}\to 0$ of the greatest interest for our purposes in which expression (37) has the form

$$\hat{f}^z_f(\mathbf{k})=\frac{i}{\hbar}\sum_{k\neq m}b_{km}(0)\mathbf{k}\cdot\mathbf{r}_{km}(0)\hat{I}^+_k\hat{I}^-_m\exp\{i\mathbf{k}\cdot\mathbf{r}_k(0)\}. \tag{38}$$

For an arbitrary time t, the values of $\hat{f}^{zQ}_f(\mathbf{k},t)$ and $\hat{f}^{zQ}_{\rm tr}(\mathbf{k},t)$ are obtained from relations (36) and (38) by the action of the ''projection evolution'' operator on them:

$$\hat{f}^{zQ}_{\rm tr}(\mathbf{k},t)\equiv\exp\{i\hat{Q}\hat{L}t\}\hat{f}^{zQ}_{\rm tr}(\mathbf{k}), \tag{39}$$

$$\hat{f}^{zQ}_f(\mathbf{k},t)\equiv\exp\{i\hat{Q}\hat{L}t\}\hat{f}^{zQ}_f(\mathbf{k}). \tag{40}$$

In order to calculate the memory matrix $M(\mathbf{k}|\mathbf{k}';\tau)$ which appears in expression (30) for the generalized Langevin equation, we need to calculate the following dynamic correlation function:

$$\langle\hat{f}^{zQ}(\mathbf{k}_1)|\hat{f}^{zQ}(\mathbf{k}_2,t)\rangle=\langle\hat{f}^{z*}(\mathbf{k}_1)\hat{f}^z(\mathbf{k}_2,t)\rangle_{\rm eq}. \tag{41}$$

In the small wave vector limit, the cross correlation contributions from the translational and flip-flop terms in expression (41) are zero. In fact, in accordance with formulas (36) and (38) these contributions are linear in the velocities and quadratic in the spin operators, i.e., we have a function which is odd with respect to the time reversal operation. Averaging these with the Gibbs equilibrium distribution function yields zero identically, which gives

$$\langle\hat{f}^{zQ}(\mathbf{k}_1)|\hat{f}^{zQ}(\mathbf{k}_2,t)\rangle=\langle\hat{f}^{z*}_{\rm tr}(\mathbf{k}_1)\hat{f}^{zQ}_{\rm tr}(\mathbf{k}_2,t)\rangle_{\rm eq}$$

$$+\langle\hat{f}^{z*}_f(\mathbf{k}_1)\hat{f}^{zQ}_f(\mathbf{k}_2,t)\rangle_{\rm eq}. \tag{42}$$

Then we are naturally confronted with the general unsolved problem involving the decoupling of the many-particle correlation functions. In cases without a clearly defined small parameter, such as the liquid-phase systems with which we are dealing, no decoupling scheme can be rigorously justified in thes present stage of development. Here we shall use the superpositional approximation, which is the simplest in mathematical terms and can indicate the type of correlations which are neglected although, like any other decoupling scheme, it cannot estimate their absolute value. This approximation generally gives reasonable quantitative results when compared with the experimental data. For instance, the equation of state for a liquid of solid spheres obtained using this approximation differs from that found in the computer ''experiments'' by no more than 20% (Ref. 25) and its application to calculate the self-diffusion coefficients of macromolecules in dilute solutions gives differences of only a few percent compared with the real experiments.[14] Thus, it is natural to first analyze our problem using the superpositional approximation.

We shall analyze the first term associated with the translational contribution in relation (42). Using relations (36) and (39), we can rewrite this in the form

$$\langle \hat{f}^z_{\rm tr}(\mathbf{k}_1)|\hat{f}^{zQ}_{\rm tr}(\mathbf{k}_2,t)\rangle = \sum_{\substack{k,m\\ \alpha,\beta}} k_\alpha k_\beta \langle V_{\alpha k}(0) V^Q_{\beta m}(t) \times \exp\{i(\mathbf{k}_2\cdot\mathbf{r}^Q_m(t) - \mathbf{k}_1\cdot\mathbf{r}_k(0))\} \times \hat{I}^{zQ}_m(t)\hat{I}^z_k\rangle_{\rm eq}, \quad (43)$$

where the superscript Q indicates the dynamic evolution of the physical quantity determined by the projection dynamics, i.e., by the evolution operator $\exp\{i\hat{Q}\hat{L}t\}$, and α and β denote the Cartesian coordinates of the vectors.

The right-hand side of formula (43) contains the many-particle dynamic correlation function which we shall subsequently calculate using the superpositional approximation:

$$\langle \hat{f}^z_{\rm tr}(\mathbf{k}_1)|\hat{f}^{zQ}_{\rm tr}(\mathbf{k}_2,t)\rangle = \sum_{\substack{k,m\\ \alpha,\beta}} k_\alpha k_\beta \langle V_{\alpha k}(0) V^Q_{\beta m}(t)\rangle_{\rm eq} \times \langle \exp\{i\mathbf{k}_2\cdot(\mathbf{r}^Q_m(t) - \mathbf{r}_m(0))\}\rangle_{\rm eq} \times \langle \exp\{i(\mathbf{k}_2\cdot\mathbf{r}^Q_m(0) - \mathbf{k}_1\cdot\mathbf{r}_k(0))\}\rangle_{\rm eq} \langle \hat{I}^{zQ}_m(t)\hat{I}^z_k\rangle_{\rm eq}. \quad (44)$$

Assuming that the system is isotropic and that the motion of the various molecules (macromolecules) is uncorrelated, we reduce the velocity–velocity correlation function to the form

$$\langle V_{\alpha k}(0) V^Q_{\beta m}(t)\rangle_{\rm eq} = \frac{1}{3}\langle \mathbf{V}^Q(t)\cdot\mathbf{V}(0)\rangle_{\rm eq}\delta_{\alpha\beta}\delta_{km}. \quad (45)$$

We substitute this relation into formula (44) and assuming that in the limit $\mathbf{k}\to 0$, the fourth factor tends to unity, we obtain

$$\langle \hat{f}^z_{\rm tr}(\mathbf{k}_1)|\hat{f}^{zQ}_{\rm tr}(\mathbf{k}_2,t)\rangle = \frac{N_s}{3}\langle \mathbf{V}^Q(t)\cdot\mathbf{V}(0)\rangle_{\rm eq} \times \langle \exp\{i(\mathbf{k}_2-\mathbf{k}_2) \cdot\mathbf{r}_m(0)\}\rangle_{\rm eq} \langle \hat{I}^{zQ}_m(t)\hat{I}^z_m\rangle_{\rm eq}, \quad (46)$$

where N_s is the number of spins in the sample.

The penultimate factor in expression (46) is the Fourier transform of the single-particle distribution function:

$$\langle \exp\{i(\mathbf{k}_2-\mathbf{k}_2)\cdot\mathbf{r}_m(0)\}\rangle_{\rm eq} = \frac{(2\pi)^3}{V}\delta(\mathbf{k}_2-\mathbf{k}_1). \quad (47)$$

We then note that in all the formulas given above we can take the spin operators $\hat{I}^z_m$ to mean their fluctuational components $\delta\hat{I}^z_m = \hat{I}^z_m - \langle \hat{I}^z_m\rangle_{\rm eq}$. This ensures that all the spin–spin dynamic correlation functions being discussed decay in the limit $t\to\infty$. We shall now define the single-spin relaxation function $P_f(t)$ by

$$\langle \hat{I}^{zQ}_m(t)\hat{I}^z_m\rangle_{\rm eq} \equiv \langle (\hat{I}^z_m)^2\rangle_{\rm eq} P_f(t) \equiv \frac{1}{3}I(I+1)P_f(t). \quad (48)$$

Note that in the general case, exact quantitative calculation of the function $P_f(t)$ is a complicated many-particle dynamic problem. However, its qualitative physical meaning is fairly clear. Damping of the correlation function $\langle \hat{I}^{zQ}_m(t)\hat{I}^z_m\rangle_{\rm eq}$ is caused by two factors: by intermolecular flip-flop processes and by spin–lattice relaxation processes. The latter are far less likely since, unlike flip-flop processes, they lead to a substantial change in the energy of the spin system. In consequence, the relaxation function $P_f(t)$ may be taken as the probability of a given spin not participating in flip-flop processes in the time t.

Thus, using formulas (47) and (48), we transform relation (46) to give

$$\langle \hat{f}^z_{\rm tr}(\mathbf{k}_1)|\hat{f}^{zQ}_{\rm tr}(\mathbf{k}_2,t)\rangle = \frac{(2\pi)^3}{3}\rho_s k^2 I(I+1) \times \langle \mathbf{V}^Q(t)\mathbf{V}(0)\rangle_{\rm eq} P_f(t)\,\delta(\mathbf{k}_2-\mathbf{k}_1). \quad (49)$$

We shall now apply similar procedures to the second term in relation (42), which represents the contribution from the torque induced by flip-flop transitions. Using relations (38) and (40), we express this in the form

$$\langle \hat{f}^z_f(\mathbf{k}_1)|\hat{f}^{zQ}_f(\mathbf{k}_2,t)\rangle = \frac{1}{\hbar^2}\sum_{\substack{k\neq m\\ k'\neq m'\\ \alpha,\alpha'}} k_\alpha k_{\alpha'} \langle b^Q_{km}(t) b_{k'm'}(0) \times r^{Q\alpha}_{km}(t) r^{\alpha'}_{k'm'}(0)\hat{I}^{+Q}_k(t)\hat{I}^-_{k'}\hat{I}^{-Q}_m(t)\hat{I}^+_{m'} \times \exp\{i(\mathbf{k}_2\cdot\mathbf{r}^Q_k(t) - \mathbf{k}_1\cdot\mathbf{r}_{k'}(0))\}\rangle_{\rm eq}, \quad (50)$$

where $r^{Q\alpha}_{km}(t)$ denotes the Cartesian coordinate α of the position vector $\mathbf{r}^Q_{km}(t)$, and $r^{\alpha'}_{k'm'}(0)$ denotes the Cartesian coordinate α' of the position vector $\mathbf{r}_{k'm'}(0)$.

Before applying the superpositional approximation to decouple the correlation function in expression (50), we represent the dynamics of the spin operators in the form

$$\hat{I}^{+Q}_k(t)\hat{I}^{-Q}_m(t) = \exp\{i\hat{Q}\hat{L}t\}\exp\{-i\hat{L}_0 t\}\exp\{+i\hat{L}_0 t\}\hat{I}^+_k\hat{I}^-_m = \exp\{i\varphi_m\}\hat{\tilde{I}}^{+Q}_k(t)\hat{\tilde{I}}^{-Q}_m(t), \quad (51)$$

where $\hat{L}_0$ is the Liouville operator generated by the Hamiltonian $\hat{H}_0 = \hat{H}_S + \hat{H}_L$,

$$\varphi_m = \gamma\mathbf{g}\cdot\int_0^t \mathbf{r}_{km}(t')dt'$$

is the phase difference between spins numbers k and m, caused by the presence of the external magnetic field gradient, and $\hat{\tilde{I}}^{\pm Q}_m$ are the spin operators in the interaction representation.

Note that formula (51) can easily be modified to allow for the magnetic nonequivalence of the different spins caused for example, by chemical shifts of the resonance frequencies. For this purpose the additional phase $\omega_{km}t$ must be added to the phase differences φ_{km}, where ω_{km} is the difference between the resonance frequencies.

An analog of the transition from formula (43) to formula (44) will involve applying the superpositional approximation for decoupling the many-particle dynamic correlation functions to the right-hand side of expression (50) in the following scheme:

$$\langle \hat{f}_f^z(\mathbf{k}_1)|\hat{f}_f^{zQ}(\mathbf{k}_2,t)\rangle = \frac{1}{\hbar^2}\sum_{\substack{k\neq m\\ k'\neq m'\\ \alpha,\alpha'}} k_\alpha k_{\alpha'}\langle b_{km}^Q(t) b_{k'm'} \times(0) r_{km}^{Q\alpha}(t) r_{k'm'}^{\alpha'}(0)\exp\{i\varphi_m(t)\}\rangle_{\rm eq} \times\langle \exp\{i(\mathbf{k}_2\cdot\mathbf{r}_k^Q(t)-\mathbf{k}_1\cdot\mathbf{r}_{k'}(0))\}\rangle_{\rm eq} \times\langle \hat{\tilde{I}}_k^{+Q}(t)\hat{I}_{k'}^-\rangle_{\rm eq}\langle \hat{\tilde{I}}_m^{-Q}(t)\hat{I}_{m'}^+\rangle_{\rm eq}. \quad (52)$$

The spin–spin correlation functions essentially represent the decay of the transverse magnetization as a result of dipole–dipole interactions. Neglecting interspin correlations, we estimate these as follows:

$$\langle \hat{\tilde{I}}_k^{+Q}(t)\hat{I}_{k'}^-\rangle = \langle \hat{\tilde{I}}_k^{-Q}(t)\hat{I}_{k'}^+\rangle = \frac{2}{3}I(I+1)\delta_{kk'}P_2(t), \quad (53)$$

where the relaxation function $P_2(t)$ describes the relaxation of the transverse components of an isolated spin in a stochastic magnetic field created by all the other spins in the system.

The situation here is exactly the same as that for the relaxation function $P_f(t)$ introduced earlier by relation (48). Strictly speaking, formula (53) is a definition of the function $P_2(t)$ whose subsequent calculation requires additional approximations based on an analysis of the physical meaning of this quantity.

The second of the spatial correlation functions in the long-wavelength approximation $\mathbf{k}_1,\mathbf{k}_2\to 0$ reduces to the single-particle static correlation function:

$$\langle \exp\{i(\mathbf{k}_2\cdot\mathbf{r}_k^Q(t)-\mathbf{k}_1\cdot\mathbf{r}_k(0))\}\rangle_{\rm eq} = \delta_{kk'}\frac{(2\pi)^3}{V}\delta(\mathbf{k}_1-\mathbf{k}_2). \quad (54)$$

Substituting relations (53) and (54) into formula (52), we obtain

$$\langle \hat{f}_f^z(\mathbf{k}_1)|\hat{f}_f^{zQ}(\mathbf{k}_2,t)\rangle = \frac{4}{9}(2\pi)^3\rho_s[I(I+1)]^2\delta(\mathbf{k}_1-\mathbf{k}_2)P_2^2(t) \times\sum_\alpha k_\alpha^2 {\sum_m}' \langle b_{km}^Q(t) r_{km}^{Q\alpha}(t) b_{km} \times(0) r_{km}^{\alpha}(0)\exp\{i\varphi_{km}(t)\}\rangle_{\rm eq}, \quad (55)$$

where the sum ${\Sigma}'_m\ldots$ indicates summation over all spins, excluding the spin labelled k, α. An average over all spins with subscript k is assumed to be included in the definition of the brackets $\langle\ldots\rangle_{\rm eq}$.

From relations (55), (49), (42), (33), (32), and the general definition of the memory matrix (16), we obtain the following expression for the density matrix contained in Eq. (30):

$$M(\mathbf{k}_1|\mathbf{k}_2;t) = \delta(\mathbf{k}_1-\mathbf{k}_2)\Bigg\{\frac{k^2}{3}\langle \mathbf{V}^Q(t)\cdot\mathbf{V}(0)\rangle P_f(t) + \frac{4}{3}[I(I+1)]^2P_2^2(t)\sum_\alpha k_\alpha^2{\sum_m}'\langle b_{km}^Q(t) r_{km}^{Q\alpha} \times(t) b_{km}(0) r_{km}^{\alpha}(0)\exp\{i\varphi_{km}(t)\}\rangle_{\rm eq}\Bigg\}. \quad (56)$$

Substituting this formula into Eq. (30), with allowance for relation (31), we obtain

$$\frac{d}{dt}\hat{\mu}^z(\mathbf{k},t) = -\sum_\alpha k_\alpha^2\int_0^t \tilde{D}_{\alpha\alpha}(\tau)\hat{\mu}^z(\mathbf{k},t-\tau)d\tau + \hat{f}^{zQ}(\mathbf{k},t), \quad (57)$$

where

$$\tilde{D}_{\rm sp}^{\alpha\alpha}(t) \equiv \tilde{D}_{\rm tr}(t) + \tilde{D}_f^{\alpha\alpha}(t), \quad (58)$$

$$\tilde{D}_{\rm tr}(t) = \frac{1}{3}\langle \mathbf{V}^Q(t)\cdot\mathbf{V}(0)\rangle P_f(t), \quad (59)$$

$$\tilde{D}_f^{\alpha\alpha}(t) = \frac{1}{12}I(I+1)\gamma^4\hbar^2P_2^2(t)\sum_m \langle L_{km}^{zzQ}(t) r_{km}^{Q\alpha}(t) \times L_{km}^{zz}(0) r_{km}^{\alpha}(0)\exp\{i\varphi_{km}(t)\}\rangle_{\rm eq}, \quad (60)$$

$$L_{km}^{zz} = \frac{1-3\cos^2\theta_{km}}{r_{km}^3}.$$

It can be seen from Eq. (57) that for long-wavelength hydrodynamic modes the characteristic relaxation time increases as k^2, whereas the characteristic decay times of the dynamic correlation functions contained in $\tilde{D}_{\rm tr}(t)$ and $\tilde{D}_f^{\alpha}(t)$ do not depend on the wave vector for $k\to\infty$. This allows us to apply the Markov approximation to Eq. (30). With allowance for relation (31), this gives

$$\frac{d}{dt}\hat{\mu}^z(\mathbf{k},t) = -\sum_\alpha k_\alpha^2 D_{\rm sp}^{\alpha\alpha}\hat{\mu}^z(\mathbf{k},t) + \hat{f}^{zQ}(\mathbf{k},t), \quad (61)$$

where

$$D_{\rm sp}^{\alpha\alpha} = \int_0^\infty \tilde{D}_{\rm sp}^{\alpha\alpha}(\tau)d\tau. \quad (62)$$

A Fourier transformation of Eq. (61) yields the diffusion equation with a Langevin source:

$$\frac{\partial}{\partial t}\hat{\mu}^z(\mathbf{r},t) = \sum_\alpha D_{\rm sp}^{\alpha\alpha}\frac{\partial^2}{\partial x_\alpha^2}\hat{\mu}^z(\mathbf{r},t) + \hat{f}^{zQ}(\mathbf{r},t). \quad (63)$$

Generally speaking, the coefficients $D_{\rm sp}^{\alpha\alpha}$ form the components of the anisotropic spin diffusion tensor. The spin diffusion tensor in accordance with relations (58) and (62) contains two contributions:

$$D_{\rm sp}^{\alpha\alpha} = D_{\rm tr} + D_f^{\alpha\alpha}, \quad (64)$$

where $D_{\rm tr}$ is the contribution associated with thermal translational displacements of the spins in space and $D_f^{\alpha\alpha}$ is the component associated with translational displacement of the spin polarization quanta as a result of flip-flop processes.

The translational contribution to the spin diffusion coefficient in accordance with formulas (59) and (62) is given by

$$D_{\rm tr} = \frac{1}{3}\int_0^\infty dt\langle \mathbf{V}^Q(t)\cdot\mathbf{V}(0)\rangle_{\rm eq}P_f(t). \quad (65)$$

Similarly for the contribution $D_f^{\alpha\alpha}$ we obtain

$$\tilde{D}_f^{\alpha\alpha}(t)=\frac{1}{12}I(I+1)\gamma^4\hbar^2\int_0^\infty dt\,P_2^2(t)$$
$$\times\sum_m\langle L_{km}^{zzQ}(t)r_{km}^{Q\alpha}(t)L_{km}^{zz}$$
$$\times(0)r_{km}^{\alpha}(0)\exp\{i\varphi_{km}(t)\}\rangle_{\rm eq}. \quad (66)$$

3. DISCUSSION OF RESULTS

Relations (63)–(66) constitute the principal result of the present study. They solve the problem of spin diffusion in condensed media in a fairly general form.

It is easy to see that the expression for $D_{\rm tr}$ resembles the Kubo–Green formula for the self-diffusion coefficient:[17,25]

$$D_{\rm sd}=\frac{1}{3}\int_0^\infty dt\langle\mathbf{V}(t)\cdot\mathbf{V}(0)\rangle_{\rm eq}. \quad (67)$$

There are two differences between them. First, the dynamic autocorrelation function $\langle\mathbf{V}^Q(t)\cdot\mathbf{V}(0)\rangle_{\rm eq}$ is determined by the ''projection'' evolution operator $\exp\{i\hat{Q}\hat{L}t\}$, whereas in relation (67) the autocorrelation function $\langle\mathbf{V}(t)\times\mathbf{V}(0)\rangle_{\rm eq}$ is determined by the complete evolution operator $\exp\{i\hat{L}t\}$. Second, unlike expression (67), the integrand of formula (65) contains the additional dynamic factor $P_f(t)$, this being the probability that the spin being studied would not participate in intermolecular flip-flop processes over the time t.

The first difference is unimportant because the operator $\hat{Q}$ projects dynamic quantities onto a half-space orthogonal to the set of spin quantities determined by relation (19), which includes any purely lattice dynamic variables. The specific heat of the spin system at the temperatures being discussed is much lower than the specific heat of the lattice, so that any kinetic processes taking place in the spin subsystem influence the lattice dynamics negligibly. This allows us to assume $\mathbf{V}^Q(t)\simeq\mathbf{V}(t)$ in relation (65), i.e., to approximate the projection dynamics with real dynamics.

The second factor $P_f(t)$ plays a fundamental role under certain circumstances. We define the average time of a flip-flop jump by

$$\tau_f\equiv\int_0^\infty dt\,P_f(t). \quad (68)$$

The autocorrelation function $\langle\mathbf{V}(t)\cdot\mathbf{V}(0)\rangle_{\rm eq}$ for times $t\gg\tau_{\max}$ decays fairly rapidly and in accordance with formula (67), the main contribution to the coefficient $D_{\rm sd}$ when integrating over time is made by times $t\sim\tau_{\max}$.

For low-molecular liquids or polymer systems with $N<N^*$ we have $\tau_{\max}\ll\tau_f$, and in relation (65) the slowly varying function $P_f(t)$ can be assumed to be close to unity for $t\ll\tau_f$. Thus, in the limit discussed relation (65) differs very little from expression (67).

Before discussing the other limiting case $\tau_{\max}\gg\tau_f$, $N>N^*$, we shall use the easily verified identity

$$\frac{1}{2}\frac{d^2}{dt^2}\langle r^2(t)\rangle_{\rm eq}=\langle\mathbf{V}(t)\cdot\mathbf{V}(0)\rangle_{\rm eq}. \quad (69)$$

We substitute this into formula (65) and after integrating by parts, we obtain

$$D_{\rm tr}=\frac{1}{6}\int_0^\infty dt\langle r^2(t)\rangle_{\rm eq}\frac{d^2}{dt^2}P_f(t). \quad (70)$$

This expression can be estimated as

$$D_{\rm tr}\simeq\frac{1}{6}\frac{\langle r^2(\tau_f)\rangle_{\rm eq}}{\tau_f}. \quad (71)$$

Thus, we obtain the first term in relation (4) which was derived earlier using the density matrix method in the second order of perturbation theory in terms of the dipole–dipole interaction operator[12] and leads to the inequality $D_{\rm sp}\geqslant D_{\rm tr}\gg D_{\rm sd}$ for polymer systems with $N>N^*$.

Note that this can be seen from formula (65). In liquid-phase systems for times $t\gg10^{-12}$ s the autocorrelation function $\langle\mathbf{V}(t)\cdot\mathbf{V}(0)\rangle_{\rm eq}$ is negative as a result of a negative correlation effect caused by inelastic backward reflection of molecules after the first collision with a nearest neighbor (see Ref. 25, for example). In polymer systems this negative long-lived tail extends to times $\tau_{\max}\simeq\tau_s N^{3.4}$ and in accordance with formula (67), it determines the self-diffusion coefficient of the macromolecule. However, for the translational contribution to the spin diffusion coefficient as given by formula (65), the contribution of the negative correlations is truncated at times of the order τ_f. Thus, we find $D_{\rm tr}\gg D_{\rm sd}$, if $\tau_f\ll\tau_{\max}$. The additional contribution associated with D_f in formula (64) merely sharpens the difference between the spin diffusion and self-diffusion coefficients in polymer systems with fairly high molecular masses.

Expression (66) is essentially the microscopic expression of the second term in formula (4) which was derived previously on the basis of phenomenological concepts. Further calculations of $D_{\rm tr}$ and D_f require a detailed analysis of various aspects of the intermolecular spin kinetics contained in $P_f(t)$ and $P_2(t)$ and of specific features of existing dynamic models of polymer systems.

One of the authors (N.F.F.) would like to thank T. N. Khazanovich and R. M. Yul'met'ev for stimulating this work, and V. D. Skirda, A. I. Maklakov, and H. Sillesscu for useful discussions. The authors thank the Volkswagen Foundation (Grant No. I/68875) and the Russian Fund for Fundamental Research (Grant No. 97-03-32668a) for financial support of this work at various stages.

*)E-mail: galina.yatsenko@ksu.ru

[1] M. Gol'dman, *Spin Temperature and Nuclear Magnetic Resonance in Solids* (Clarendon Press, Oxford, 1970) [Mir, Moscow, 1972].
[2] A. Abragam, *The Principles of Nuclear Magnetism* (Clarendon Press, Oxford, 1961) [Moscow, 1963].
[3] I. V. Aleksandrov, *Theory of Magnetic Relaxation* [in Russian], Nauka, Moscow (1975).
[4] G. R. Khutsishvili, Usp. Fiz. Nauk **87**, 211 (1965) [Sov. Phys. Usp. **8**, 734 (1965)].
[5] H.-W. Spiess, Annu. Rev. Mater. Sci. **21**, 131 (1991).
[6] D. E. Demco, A. Johansson, and J. Tegenfeld, Solid State Nucl. Magn. Reson. **4**, 13 (1995).
[7] A. I. Maklakov, V. D. Skirda, and N. F. Fatkullin, in *Encyclopedia of Fluid Mechanics, Vol. 9, Polymer Flow Engineering*, edited by N. P. Cheremisinoff (Gulf Publishing, Houston, 1990), Chap. 22, p. 705.

[8] P. T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy* (Clarendon Press, Oxford, 1991).
[9] R. Kimmich, *NMR: Tomography, Diffusometry, Relaxometry* (Springer, Berlin, 1997).
[10] N. F. Fatkullin, in Abstracts of Papers presented at the Ninth Ampere Summer School, Novosibirsk (1987), p. 209.
[11] N. F. Fatkullin, in Abstracts of Papers presented at the Ninth Specialized AMPERE Colloquium on Magnetic Resonance in Polymers, Prague (1989), p. 91.
[12] N. F. Fatkullin, Zh. Éksp. Teor. Fiz. **99**, 1013 (1991) [Sov. Phys. JETP **72**, 563 (1991)].
[13] R. Kimmich, G. Schur, and M. Kopf, in *Progress in NMR Spectroscopy*, (1987), Vol. 20, p. 385.
[14] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1989).
[15] K. S. Schweizer, J. Chem. Phys. **91**, 5802 (1989).
[16] N. Fatkullin and R. Kimmich, J. Chem. Phys. **101**, 822 (1994).
[17] D. Forster, *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions*, Benjamin, Reading, Mass. (1975) [Atomizdat, Moscow (1980)].
[18] V. I. Sundukov, V. D. Skirda, A. I. Maklakov, and M. M. Doroginitskiĭ, Vysokomol. Soedin., Ser. B **28**, 382 (1986).
[19] V. D. Skirda, M. M. Doroginizkii, V. I. Sundukov *et al.*, Macromol. Rapid Commun. **9**, 603 (1988).
[20] I. Chang, F. Fujara, and B. Geil *et al.*, J. Non-Cryst. Solids **172–174**, 674 (1994).
[21] E. Fischer, R. Kimmich, and N. Fatkullin, J. Chem. Phys. **106**, 9883 (1997).
[22] M. E. Komlush, P. T. Callaghan, in Abstracts of Papers presented at the Fourth International Conference on Magnetic Resonance Microscopy and Macroscopy, Albuquerque, NM (1997), p. 89.
[23] C. H. Wang, *Spectroscopy of Condensed Media* (Academic Press, Orlando, 1985).
[24] B. J. Berue and R. Pecora, *Dynamic Light Scattering* (Wiley, New York, 1976).
[25] R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics*, Vols. 1 and 2 (Wiley, New York 1975) [Mir, Moscow, 1978].

Translated by R. M. Durham

Inelastic scattering of phonons by quadrupole defects with internal degrees of freedom

T. N. Antsygina, V. A. Slyusarev, and K. A. Chishko*)

B. I. Verkin Low-Temperature Physicotechnical Institute, National Academy of Sciences, 310164 Kharkov, Ukraine

(Submitted 19 August 1997)
Zh. Éksp. Teor. Fiz. **114**, 555–569 (August 1998)

We solve the quantum mechanical problem of the inelastic scattering of phonons by a quadrupole defect in a crystal lattice for the case of solid parahydrogen whose matrix contains pair complexes of H_2 orthomolecules. By employing the pseudospin approximation for the operator of the energy of quadrupole–quadrupole interaction of the molecules in an orthopair we derive an effective Hamiltonian that describes the interaction of phonons with a pair quadrupole orthodefect in the lattice. We set up the scattering matrix and calculate the effective phonon relaxation time $\tau(\omega,T)$ as a function of the frequency ω and the crystal temperature T. We also find that a pair quadrupole defect, which has a complicated system of levels, can be replaced by an effective two-level system with temperature-dependent parameters. The fact that a pair quadrupole orthocluster has internal degrees of freedom results in a resonant scattering peak near a certain critical temperature T_0. Our estimates for H_2 yield $T_0 \simeq 6$–7 K. Finally, we discuss the contribution of this mechanism to the low-temperature thermal conductivity of solid hydrogen. © *1998 American Institute of Physics.* [S1063-7761(98)01408-5]

1. INTRODUCTION

The study of mechanisms of phonon scattering in crystals constitutes one of the central problems of solid state physics.[1,2] The main channel of relaxation of the phonon subsystem in a nonconducting medium is the phonon–phonon interaction.[1,2] This channel includes normal processes (or *N*-processes) and umklapp processes (or *U*-processes), whose properties have been thoroughly studied.[2] There is, however, a broad class of relaxation phenomena that occur because of the interaction of phonons with the structure defects and impurities in the crystal lattice.[3] Scattering by point centers (vacancies and interstitial and substitutional impurities), linear defects (dislocations), and planar defects (grain boundaries, stacking faults, and interphase boundaries) has a strong effect on the thermal conductivity of the samples, especially in the low-temperature range where the intensity of phonon–phonon processes is not very high.

All aspects of the kinetics of the phonon subsystem become especially important when we study atomic and molecular cryocrystals (solidified inert gases and solid hydrogen, oxygen, nitrogen, etc.) and solid solutions based on these cryocrystals.[4] Most interesting in this series are the hydrogen and helium crystals, since their observed physical properties are a consequence of essentially quantum effects, which do not manifest themselves in classical solids. As an example directly related to the topic of the present paper we would like to mention the anomalous temperature behavior of the thermal conductivity coefficient $\kappa(T)$ of solid H_2 in the 1–7 K temperature range.[5–8] The curve representing this behavior has a peak at $\simeq 4$–5 K, whose nature and also behavior $\kappa(T) \propto T^m (m \simeq 2$–$3)$ at temperatures below 5 K are still topics of discussions.

Hydrogen crystals are formed by diatomic molecules each of which is either in the ortho modification (o-H_2) or in the para modification (p-H_2). To any finite temperature of the crystal there corresponds a certain equilibrium concentration x of the ortho modification. However, in low-temperature experiments one can prepare H_2 crystals with any ortho–para composition desired, since in the absence of external perturbations the ortho–para conversion times are extremely long.

Bohn and Mate[5] measured the thermal conductivity of solid H_2 at low orthocomponent concentrations and found that the height of the low-temperature peak in the κ vs. T curve decreases with increasing x. Further experiments carried out by Huebler and Bohn[6] and Reynolds and Anderson[7] showed that such behavior also holds for samples with high o-H_2 content. Bohn and Mate[5] suggested semiempirical κ vs. T curves, which fit fairly well the experimental data in the low-temperature range if one assumes that the main phonon relaxation mechanism in this range is the scattering by grain boundaries in polycrystalline hydrogen and the average crystallite size in the sample decreases with increasing x. Within this setting, however, the crystallite size is present only as a fitting parameter, since the literature contains no data on the correlation between grain size and orthocomponent concentration in H_2 crystals.

Ebner and Sung[9] attempted to theoretically explain the anomalous thermal conductivity of H_2 by allowing for phonon relaxation at single orthomolecules is the paramatrix. They found that the contribution of this mechanism to the total thermal conductivity is moderate and that the mechanism cannot be used to interpret the temperature dependence of κ unambiguously. This comes as no surprise, however, since the interpretation of an orthomolecule as an impurity in the paramatrix of solid hydrogen is largely a matter of taste. An orthoimpurity is really not a structure defect, since an isolated orthomolecule does not differ from the surrounding

paramolecules either in intermolecular interaction parameters or in mass. A distinctive feature of an orthomolecule is the finite nuclear spin and quadrupole electric moment. On the other hand, since the intermolecular distances in the H_2 crystal are large compared to the size of the molecules, an orthomolecule can rotate almost freely at the site it occupies. Estimates that use the well-known parameters of the H_2 molecule[4] show that the ''angular velocity'' of such rotation exceeds the Debye frequency ω_D of hydrogen by a factor of 1.5. Thus, from the viewpoint of phonon processes, an orthoimpurity is spherically symmetric and spatially isotropic, i.e., is practically a paramolecule.

At the same time it is known[10,11] that an significant number of orthomolecules are present in the parahydrogen matrix in the form of double, triple, etc. clusters, which means that we must establish their role in bringing about the experimentally observable thermal conductivity of solid H_2. This problem appears to be even more natural if we note that, as shown by Nakamura,[12] the contribution of pair orthoclusters can explain the low-temperature anomalies of the thermal conductivity of solid ortho–para mixtures. A pair of orthomolecules occupying two neighboring sites of the matrix comprise a pair cluster with its own energy spectrum, which differs from the rotational spectrum of a single molecule. As a result, one can expect resonant phonon scattering in the range of temperatures comparable to the distance between the levels in an orthopair. It is natural to assume that in its main features this mechanism is similar to the resonant scattering of phonons by the quasilocal levels of heavy atomic impurities in p-H_2 (see Refs. 13 and 14).

Kokshenev[15] attempted to develop a theory of thermal conductivity of solid hydrogen that takes into account the effects of inelastic scattering of phonons by pair orthocomplexes from the start. The theory was then applied to the corresponding experimental data.[8] Although there is some quantitative agreement with the experimental data at 1–3 K (the theory developed in Ref. 15 neither reproduces nor explains the peak in thermal conductivity at $T\simeq 5$ K), the corresponding calculation is not very consistent. In particular, the contribution of orthopairs to the total phonon relaxation time was calculated in Ref. 15 with $\langle \tau_p^{-1}\rangle$ unjustifiably replaced by $\langle \tau_p\rangle^{-1}$ before the kinetic equation had even been solved. After that $\langle \tau_p\rangle$ was arbitrarily increased by a factor of 5.4 to achieve quantitative agreement with the experimental data. In addition to this, the total thermal conductivity took into account the scattering by crystallite boundaries and by single orthomolecules, with the parameters pertaining to these mechanisms considered fitting parameters. The outcome of all this is that the results of Ref. 15 cannot be used to arrive at definite conclusions concerning the relative role of orthopairs in the formation of the low-temperature thermal conductivity of solid H_2.

Thus, we see that providing a consistent description of the interaction of phonons and quadrupole clusters in cryocrystal matrices is still an important task. The aim of this paper is to give a rigorous solution of the quantum mechanical problem of the inelastic scattering of phonons by a pair orthocomplex in the parahydrogen matrix.

2. STATEMENT OF THE PROBLEM

We can describe the anisotropic interaction of two orthohydrogen molecules with a high accuracy by the quadrupole part of the interaction.[4] The corresponding Hamiltonian is

$$V_Q=\frac{3}{4}\frac{Q^2}{R^5}(35Q_1^{\alpha\beta}Q_2^{\gamma\delta}n_\alpha n_\beta n_\gamma n_\delta - 20Q_1^{\alpha\beta}Q_2^{\beta\gamma}n_\alpha n_\gamma+2Q_1^{\alpha\beta}Q_2^{\beta\alpha}), \tag{1}$$

where Q is the electric quadrupole moment of the orthomolecule, $\mathbf{R}$ is the distance between the two impurity molecules, $\mathbf{n}=\mathbf{R}/R$ is the unit vector along the line connecting the centers of the molecules, and $Q_j^{\alpha\beta}$ is the operator of the dimensionless quadrupole moment ($j=1,2$ and $\alpha,\beta=x,y,z$). Since the o-H_2 molecule is usually in a state with the orbital quantum number $l=1$ (but with three possible values of the magnetic quantum number $m=-1,0,1$), it is convenient to write the operator $Q_j^{\alpha\beta}$ in the pseudospin approximation,[12] in which

$$Q_j^{\alpha\beta}=-\tfrac{1}{5}(S_j^\alpha S_j^\beta+S_j^\beta S_j^\alpha-\tfrac{4}{3}\Delta_{\alpha\beta}), \tag{2}$$

where $\mathbf{S}_j$ is the spin operator with $S=1$.

Below we will be interested in the interaction of the phonon subsystem and a pair defect. The Hamiltonian of the corresponding interaction can be obtained by the following reasoning. Let $\mathbf{R}_0$ be the equilibrium distance between the centers of the two orthomolecules ($\mathbf{n}_0=\mathbf{R}_0/R_0$) positioned at two neighboring lattice sites. Assuming $\mathbf{R}=\mathbf{R}_0+\mathbf{u}$, where $\mathbf{u}=\mathbf{u}_1-\mathbf{u}_2$ is the relative displacement of the two o-H_2 molecules from their positions of equilibrium, we expand the Hamiltonian (1) in powers of the small displacement $\mathbf{u}$. Here we allow for the fact that a variation of R in the denominator of (1) does not give rise to off-diagonal elements describing transitions between the various levels of the subsystem consisting of two orthomolecules. We ignore this variation, since it can yield only a correction to elastic phonon scattering and is not involved in the inelastic effects we are interested in. Thus, we should perform the expansion only for the unit vector $\mathbf{n}$. As a result $\mathbf{n}$ acquires an increment $\delta\mathbf{n}_0$ equal to

$$\delta\mathbf{n}_0=\boldsymbol{\varphi}\times\mathbf{n}_0,\quad \boldsymbol{\varphi}=\frac{1}{R_0}\mathbf{n}_0\times\mathbf{u}. \tag{3}$$

This increment is a rotation of the system as whole through the angle φ about the axis directed along the vector $\boldsymbol{\varphi}$. The structure of the Hamiltonian (1) will be preserved if we rotate the vectors $\mathbf{S}_j$ through the same angle. Indeed, the rotation (3) corresponds to the operator of a unitary transformation to which the vectors $\mathbf{S}_j$ are subjected,

$$U=\exp(-i\boldsymbol{\varphi}\cdot\mathbf{S}), \tag{4}$$

where $\mathbf{S}=\mathbf{S}_1+\mathbf{S}_2$. Since $\mathbf{S}_j$ enters into (1) either in the combination $\mathbf{S}_i\cdot\mathbf{S}_j$ (invariant under transformation (4)) or in the combination $\mathbf{S}_j\cdot\mathbf{n}$ (transformed into $\mathbf{S}_j\cdot\mathbf{n}_0$), the Hamiltonian (1) of quadrupole interaction proves to be invariant under the transformation (4). The physical meaning of this result is that rotations (and translations) of a molecule as a whole cannot excite the internal degrees of freedom. What the transforma-

tion (4) does is separate explicitly the terms in the Hamiltonian that are responsible for excitation of internal degrees of freedom and lead to inelastic scattering of phonons by pair orthoclusters. A similar transformation was proposed by Dyson[16] for solving problems of quantum field theory.

Thus, at first glance it would appear that within this approximation no essentially inelastic effects manifest themselves. However, we now turn our attention to the fact that the operator (4) does not commute with the part of the kinetic-energy operator that describes the translational motion of the orthomolecules:

$$H_{\rm kin}=\frac{1}{2M}(\hat{\mathbf{p}}_1^2+\hat{\mathbf{p}}_2^2), \tag{5}$$

where the $\hat{\mathbf{p}}_i$ are the operators of the momenta of the molecules, and M is the molecule's mass. In the final analysis it is the noncommutativity of the operator (5) and the operator (4), which diagonalizes the potential energy (1), that guarantees the occurrence of off-diagonal transitions. Applying the transformation (4) to (5) and allowing for the fact that

$$U^{\dagger}\hat{\mathbf{p}}_j^2U=\left(1-i\frac{\mathbf{u}_j}{R_0}\cdot(\mathbf{n}_0\times\mathbf{S})\right)\hat{\mathbf{p}}_j^2\left(1+i\frac{\mathbf{u}_j}{R_0}\cdot(\mathbf{n}_0\times\mathbf{S})\right)$$
$$=\hat{\mathbf{p}}_j^2+\frac{2}{R_0}\hat{\mathbf{p}}_j\cdot(\mathbf{n}_0\times\mathbf{S}),$$

we find that

$$H_{\rm kin}\to\tilde{H}_{\rm kin}=U^{\dagger}H_{\rm kin}U=H_{\rm kin}+\Delta H_{\rm kin}, \tag{6}$$

where

$$\Delta H_{\rm kin}=\frac{1}{MR_0}(\mathbf{n}_0\times\mathbf{S})(\hat{\mathbf{p}}_1-\hat{\mathbf{p}}_2).$$

We can assume, without loss of generality, that one of the molecules ($j=1$) of orthohydrogen that form the orthopair is at the origin of the coordinate system. Then, in the second quantization representation for the phonon variables, the momentum operators $\hat{\mathbf{p}}_j$ have the form

$$\hat{\mathbf{p}}_1=\frac{1}{\sqrt{2N}}\sum_{\mathbf{k},\alpha}\sqrt{M\omega_{\mathbf{k}\alpha}}\,\boldsymbol{\varepsilon}_{\alpha}(\mathbf{k})(a_{\mathbf{k}\alpha}^{\dagger}+a_{-\mathbf{k}\alpha}), \tag{7}$$

$$\hat{\mathbf{p}}_2=\frac{1}{\sqrt{2N}}\sum_{\mathbf{k},\alpha}\sqrt{M\omega_{\mathbf{k}\alpha}}\,\boldsymbol{\varepsilon}_{\alpha}(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{n}_0R_0)$$
$$\times(a_{\mathbf{k}\alpha}^{\dagger}+a_{-\mathbf{k}\alpha}). \tag{8}$$

Here N is the number of sites in the lattice, $\boldsymbol{\varepsilon}_{\alpha}(\mathbf{k})$ are the polarization vectors of the phonons, satisfying the condition $\boldsymbol{\varepsilon}_{\alpha}^{*}(-\mathbf{k})=\boldsymbol{\varepsilon}_{\alpha}(\mathbf{k})$, and $\omega_{\mathbf{k}\alpha}$ is the frequency of a phonon with momentum $\mathbf{k}$ and polarization α, and $a_{\mathbf{k}\alpha}^{\dagger}$ and $a_{\mathbf{k}\alpha}$ are the phonon creation and annihilation operators. Next it is convenient to write the operator $\mathbf{S}$ as

$$\mathbf{S}=\mathbf{n}_0S^z+\mathbf{e}_1S^x+\mathbf{e}_2S^y, \tag{9}$$

where the unit vectors $\mathbf{n}_0$ and $\mathbf{e}_j$ satisfy the following conditions:

$$\mathbf{e}_1\times\mathbf{e}_2=\mathbf{n}_0,\quad \mathbf{n}_0\times\mathbf{e}_1=\mathbf{e}_2,\quad \mathbf{e}_2\times\mathbf{n}_0=\mathbf{e}_1.$$

Allowing for the fact that an orthohydrogen molecule has the same mass as a parahydrogen molecule and differs only in the fact that it has an intrinsic angular momentum, we can use Eqs. (6)–(9) and write the total Hamiltonian of the system as

$$H=H_0+\Delta H_{\rm kin},\quad H_0=H_{\rm ph}+V_Q,$$

$$H_{\rm ph}=\sum_{\mathbf{k},\alpha}\omega_{\mathbf{k}\alpha}(a_{\mathbf{k}\alpha}^{\dagger}a_{\mathbf{k}\alpha}+\tfrac{1}{2}), \tag{10}$$

$$\Delta H_{\rm kin}=\frac{1}{\sqrt{N}}\sum_{\mathbf{k},\alpha}V_{\mathbf{k}\alpha}(a_{\mathbf{k}\alpha}^{\dagger}+a_{-\mathbf{k}\alpha}),$$

where we have introduced the notation

$$V_{\mathbf{k}\alpha}=R_{\alpha}(\mathbf{k})S^{+}-R_{\alpha}^{*}(-\mathbf{k})S^{-}, \tag{11}$$

where

$$R^{\alpha}(\mathbf{k})=\frac{1}{2}\sqrt{\frac{\omega_{\mathbf{k}\alpha}}{M}}(\mathbf{k}\cdot\mathbf{n}_0)(\boldsymbol{\varepsilon}_{\alpha}(\mathbf{k})\cdot\mathbf{e}^{*}), \tag{12}$$

$$S^{\pm}=S^x\pm iS^y,\quad \mathbf{e}=\frac{1}{\sqrt{2}}(\mathbf{e}_1+i\mathbf{e}_2).$$

In writing (11) we used the long-wavelength approximation, since a specific feature of quantum crystals is that their Debye temperature Θ_D is much higher than their melting point. Thus, we are always within the low-temperature range $T\ll\Theta_D$.

3. THE MATRIX OF PHONON SCATTERING BY A QUADRUPOLE DEFECT

Let us write the equations of motion for the operator $a_{\mathbf{k}\alpha}$. Since the interaction between the pseudospin and phonon subsystems is weak, the effect of one subsystem on the other can be taken into account in the mean-field approximation. Then the equation of motion has the form

$$i\frac{d}{dt}a_{\mathbf{k}\alpha}=\omega_{\mathbf{k}\alpha}a_{\mathbf{k}\alpha}+\frac{1}{\sqrt{N}}\langle V_{\mathbf{k}\alpha}\rangle. \tag{13}$$

Here the angle brackets indicate averaging, $\langle\cdots\rangle=\mathrm{Tr}\{\rho(t)\cdots\}$, with the statistical operator $\rho(t)$, which satisfies the equation $i(d\rho/dt)=[H,\rho(t)]$ and the initial condition

$$\rho(t)|_{t=-\infty}=\rho_0=\frac{\exp(-H_0/T)}{\mathrm{Tr}\exp(-H_0/T)}. \tag{14}$$

The initial condition means that at $t=-\infty$ the system was in a state of statistical equilibrium and was described by a Gibbs canonical distribution.

The average value $\langle V_{\mathbf{k}\alpha}(t)\rangle$ to first order in $\Delta H_{\rm kin}$ is given by the following relationship:[17]

$$\langle V_{\mathbf{k}\alpha}(t)\rangle=\langle V_{\mathbf{k}\alpha}\rangle_0-i\int_{-\infty}^{t}d\tau\langle[V_{\mathbf{k}\alpha}(t),\Delta H_{\rm kin}(\tau)]\rangle_0, \tag{15}$$

where $\langle\cdots\rangle_0$ indicates averaging with the equilibrium distribution function (14). Extending the integral in (15) over all values of τ and allowing for the fact that in our case we have $\langle V_{\mathbf{k}\alpha}\rangle_0=0$, we obtain

$$\langle V_{\mathbf{k}\alpha}(t)\rangle=\int_{-\infty}^{\infty}d\tau\langle\langle V_{\mathbf{k}\alpha}(t)|\Delta H_{\rm kin}(\tau)\rangle\rangle, \tag{16}$$

where by $\langle\langle A(t)|B(t')\rangle\rangle$ we denote the retarded two-time Green's function. Plugging (16) into (13) and using (10), we arrive at the following equation for $a_{\mathbf{k}\alpha}$:

$$i\frac{d}{dt}a_{\mathbf{k}\alpha}=\omega_{\mathbf{k}\alpha}a_{\mathbf{k}\alpha}+\frac{1}{N}\sum_{\mathbf{p},\gamma}\int_{-\infty}^{\infty}d\tau\langle\langle V_{\mathbf{k}\alpha}(t)|V_{\mathbf{p}\gamma}(\tau)\rangle\rangle \times[a^{\dagger}_{\mathbf{p}\gamma}(\tau)+a_{-\mathbf{p}\gamma}(\tau)]. \tag{17}$$

Let us apply the Fourier transformation with respect to time to (17). Then, bearing in mind (11), for the Fourier transforms $A_{\mathbf{k}\alpha}$ of the operators $a_{\mathbf{k}\alpha}$ we obtain

$$A_{\mathbf{k}\alpha}=\Delta_{\mathbf{k}\mathbf{q}}\Delta_{\alpha\beta}+\frac{2\pi}{\omega_{\mathbf{q}\beta}-\omega_{\mathbf{k}\alpha}}\frac{1}{N}\sum_{\mathbf{p},\gamma}\{R_{\alpha}(\mathbf{k})R^{*}_{\gamma}(-\mathbf{p}) \times\langle\langle S^{+}|S^{-}\rangle\rangle_{\omega_{\mathbf{q}\beta}}+R^{*}_{\alpha}(-\mathbf{k})R_{\gamma}(\mathbf{p}) \times\langle\langle S^{-}|S^{+}\rangle\rangle_{\omega_{\mathbf{q}\beta}}\}(A^{\dagger}_{\mathbf{p}\gamma}+A_{-\mathbf{p}\gamma}). \tag{18}$$

The Green's function in (18) can be calculated by the standard method:[17]

$$\langle\langle S^{-}|S^{+}\rangle\rangle_{\Omega}\equiv\frac{P(\Omega)}{2\pi}=\frac{1}{2\pi}\sum_{i,j}|S^{-}_{ij}|^{2}\frac{\nu_i-\nu_j}{\Omega-E_i+E_j+i\delta}. \tag{19}$$

Here $S^{\pm}_{ij}=(\psi^{*}_i S^{\pm}\psi_j)$, where ψ_i and E_i are the eigenfunctions and eigenvectors of the operator V_Q, and $\nu_i=\exp(-E_i/T)/Z_0$, with $Z_0=\Sigma_i\exp(-E_i/T)$. Similarly, $\langle\langle S^{+}|S^{-}\rangle\rangle_{\Omega}=P(-\Omega)/2\pi$. Direct calculations show that the function $P(\Omega)$ is even and reduces to

$$P(\Omega)=\sum_{i,j}|S^{-}_{ij}|^{2}\frac{(\nu_i-\nu_j)(E_i-E_j)}{(E_i-E_j)^{2}-\Omega^{2}}. \tag{20}$$

If we use (19), Eq. (18) becomes

$$A_{\mathbf{k}\alpha}=\Delta_{\mathbf{k}\mathbf{q}}\Delta_{\alpha\beta}+\frac{P(\omega_{\mathbf{q}\beta})}{\omega_{\mathbf{q}\beta}-\omega_{\mathbf{k}\alpha}}\frac{1}{N}\sum_{\mathbf{p},\gamma}[R^{*}_{\alpha}(-\mathbf{k})R_{\gamma}(\mathbf{p})+R_{\alpha}(\mathbf{k}) \times R^{*}_{\gamma}(-\mathbf{p})](A^{\dagger}_{\mathbf{p}\gamma}+A_{-\mathbf{p}\gamma}). \tag{21}$$

Similarly,

$$A^{\dagger}_{\mathbf{k}\alpha}=-\frac{P(\omega_{\mathbf{q}\beta})}{\omega_{\mathbf{q}\beta}+\omega_{\mathbf{k}\alpha}}\frac{1}{N}\sum_{\mathbf{p},\gamma}[R^{*}_{\alpha}(\mathbf{k})R_{\gamma}(\mathbf{p})+R_{\alpha}(-\mathbf{k}) \times R^{*}_{\gamma}(-\mathbf{p})](A^{\dagger}_{\mathbf{p}\gamma}+A_{-\mathbf{p}\gamma}). \tag{22}$$

The system of linear integral equations (21) and (22) with degenerate kernels can be solved by the standard method.[18] For simplicity we ignore the difference between longitudinal and transverse phonons, i.e., we assume $\omega_{k\alpha}\equiv\omega_k=sk$, where $s^{-3}=\frac{1}{3}(s_l^{-3}+2s_t^{-3})$, with s_l and s_t the longitudinal and transverse speeds of sound. As a result we obtain

$$A_{\mathbf{k}\alpha}=\Delta_{\mathbf{k}\mathbf{q}}\Delta_{\alpha\beta}+\frac{1}{N}\frac{T^{\alpha\beta}_{\mathbf{k}\mathbf{q}}}{\omega_q-\omega_k+i\delta}. \tag{23}$$

Here the T matrix has the form

$$T^{\alpha\beta}_{\mathbf{k}\mathbf{q}}=\frac{1}{N}\frac{P(\omega_q)}{1-B(\omega_q)}[R_{\alpha}(\mathbf{k})R^{*}_{\beta}(\mathbf{q})+R^{*}_{\alpha}(-\mathbf{k})R_{\beta}(-\mathbf{q})], \tag{24}$$

where

$$B(\omega_q)=\frac{P(\omega_q)}{2MN}\sum_{\mathbf{p}}\frac{\omega_p^{2}(\mathbf{p}\cdot\mathbf{n}_0)^{2}}{\omega_q^{2}-\omega_p^{2}}. \tag{25}$$

We now average (25) over $\mathbf{n}_0$ and replace summation over $\mathbf{p}$ by integration:

$$\frac{1}{N}\sum_{\mathbf{p}}f(p)=\frac{3}{k_D^{2}}\int_0^{k_D}dp\,p^{2}f(p),$$

where k_D is the Debye wave vector. As a result we have

$$B(\omega)=\frac{P(\omega)}{2Mk_D^{3}}\int_0^{k_D}dp\,p^{4}\frac{\omega_p^{2}}{\omega^{2}-(\omega_p+i\delta)^{2}}. \tag{26}$$

Integrating with respect to $\omega_p=sp$ and using the identity

$$\frac{1}{E-\omega\pm i\delta}=\mathscr{P}\frac{1}{E-\omega}\mp i\pi\delta(E-\omega),\quad \delta\to+0$$

(here $\mathscr{P}$ is the principal-value symbol), we finally obtain

$$B(\omega)=-\frac{\gamma\omega_D P(\omega)}{4}\left[\frac{2}{5}+i\pi\left(\frac{\omega}{\omega_D}\right)^{5}\right],\quad \gamma=\frac{\omega_D}{Ms^{2}}, \tag{27}$$

where $\omega_D=sk_D$ is the Debye frequency. In calculating the principal value of the integral with respect to ω_p we ignored ω^2 in the denominator in (26), since ω is small compared to ω_D.

The T matrix (24) is linked to the effective relaxation time $\tau_p(k)$ of phonons in pair complexes averaged over phonon polarizations through the relationship

$$\frac{1}{\tau_p(k)}=-\frac{2n_p}{3}\sum_{\alpha}\mathrm{Im}\,T^{\alpha\alpha}_{\mathbf{k}\mathbf{k}}, \tag{28}$$

where n_p is the orthopair concentration. Using (24) and (12) and averaging (28) over $\mathbf{n}_0$, we finally get

$$\frac{1}{\tau_p(\omega)}=-\frac{n_p\omega^{3}}{9Ms^{2}}\,\mathrm{Im}\,\frac{P(\omega)}{1-B(\omega)}. \tag{29}$$

To keep the notation concise we have put $\omega_k\equiv\omega$. Plugging (27) into (29), we arrive at the following expression for the relaxation time:

$$\frac{1}{\tau_p(\omega)}=-\frac{\gamma n_p\omega_D^{2}}{9}\left(\frac{\omega}{\omega_D}\right)^{3} \times\mathrm{Im}\left\{\frac{1}{P(\omega)}+\gamma\omega_D\left[\frac{2}{5}+i\pi\left(\frac{\omega}{\omega_D}\right)^{5}\right]\right\}^{-1}. \tag{30}$$

The quantity $P(\omega)$ in (30) has a complicated structure (see Eq. (20)). At $\omega=0$ Eq. (20) becomes

$$P(0)=\sum_{i,j}|S_{ij}^{-}|^2\frac{\nu_i-\nu_j}{E_i-E_j}\equiv-\frac{\sigma_1(T)}{\omega_D},\tag{31}$$

while in the limit $\omega\to\infty$,

$$\omega^2P(\omega)\to-\sum_{i,j}|S_{ij}^{-}|^2(\nu_i-\nu_j)(E_i-E_j)\equiv\omega_D\sigma_2(T).\tag{32}$$

By direct computer calculations it can be shown that in the intermediate range of ω-values the function $P(\omega)$ is approximated fairly well by the expression

$$P(\omega)=C\frac{\omega_0}{\omega^2-\omega_0^2},\tag{33}$$

where $C=\sqrt{\sigma_1\sigma_2}$, and $\omega_0^2=\omega_D^2\sigma_2/\sigma_1$.

Actually this approximation corresponds to a situation in which the orthopair, which has a complicated system of levels, can be replaced by an effective two-level system whose parameters, however, are functions of the temperature. Using (33), we arrive at the final expression for the reciprocal relaxation time:

$$\frac{1}{\tau_p(u)}=\frac{4n_p\omega_D}{9\pi}\frac{\xi^2u^8}{(u^2-bu_0^2)^2+\xi^2u^{10}}.\tag{34}$$

Here we have introduced the following notation:

$$u=\frac{\omega}{\omega_D},\quad u_0=\frac{\omega_0}{\omega_D},\quad \xi=\frac{\pi\gamma}{4}\sigma_2,\quad b=1-0.1\gamma\sigma_1.$$

Thus, the expression for the reciprocal time of the phonon scattering by a pair orthocluster is very similar in structure to the corresponding expression for the reciprocal time of phonon scattering by a heavy impurity.[19] At the same time, there are important differences, the principal of which is that the ''resonant frequency'' and the parameter ξ in (34) are temperature-dependent and the ratio ω/ω_D in (34) is raised to a high power.

The dependence of the relaxation time τ_p on the dimensionless frequency u is determined by the relationship between three parameters, γ, $\sigma_1(T)$, and $\sigma_2(T)$, with the last two being temperature-dependent. Using Eqs. (31) and (32) and the explicit expressions for the eigenfunctions and eigenvalues of the operator (1) (see Refs. 12 and 20), we can express $\sigma_1(T)$ and $\sigma_2(T)$ as follows:

$$\sigma_1(T)=\frac{2}{5}\frac{\omega_D}{\Gamma_0}f_1(\beta),\tag{35}$$

$$f_1(\beta)=\frac{e^{4\beta}}{Z_0}(1-e^{-5\beta})(7+3e^{-5\beta}),$$

$$\sigma_2(T)=40\frac{\Gamma_0}{\omega_D}f_2(\beta),\tag{36}$$

$$f_2(\beta)=\frac{e^{4\beta}}{Z_0}(1-e^{-5\beta})(4+3e^{-5\beta}),$$

where $\Gamma_0=(6/25)(Q^2/R_0^5)$, $\beta=\Gamma_0/T$, and the orthopair partition function introduced earlier is $Z_0=4+2e^{4\beta}+2e^{-\beta}+e^{-6\beta}$. The quantity $b(T)$ in (34) may be either positive or negative, depending on the temperature. Clearly, as $T\to\infty$, the functions f_1 and f_2 tend to zero, with $u_0^2\to$const and $b\to1$. Thus, at high temperatures the function $\tau_p^{-1}(u)$ has a resonant peak at the frequency $u_r=u_0\sqrt{b}$. The peak's height and dimensionless width are, respectively,

$$\tau_{pm}^{-1}=\frac{4n_p\omega_D}{9\pi u_r^2},\quad \delta_m\approx\frac{1}{2}\xi u_r^4.\tag{37}$$

In the limit $T\to0$, the functions $f_1(\beta)$ and $f_2(\beta)$ (and hence u_0^2) tend to finite values and b becomes $b(0)=1-0.14\gamma(\omega_D/\Gamma_0)$. Hence, the sign of $b(0)$ is determined by the parameters of the specific system. Usually we have $\omega_D\gg\Gamma_0, Ms^2$ in cryocrystals, so that at low temperatures b is negative and there is no resonant scattering.

Thus, we can say that resonant scattering of phonons by a pair orthocluster occurs at temperatures higher than a certain critical value T_0, which can be found by solving the transcendental equation $b(T_0)=0$ or, which is the same,

$$\sigma_1(T_0)=\frac{10}{\gamma}.\tag{38}$$

As T_0 is approached from higher temperatures, both u_r and the width δ_m of the resonance peak tend to zero, while the height of the peak increases without limit. Thus, the function (34) has a singularity at $u=u_r$ only at a single value of the parameters $T=T_0$. By comparison, the function $\tau_p^{-1}(\omega)$, determined earlier in Ref. 15, is essentially singular at all temperatures: it has temperature-independent delta-function singularities at frequencies $\omega=E_i-E_j$. Using the values $\Gamma_0\sim1$ K and $\omega_d\simeq117$ K for solid hydrogen (known from the literature[5]), we find that $T_0\simeq6$–7 K. This result has an obvious physical meaning: resonant scattering is important only when the average phonon energy becomes comparable to the separation of levels in the orthoimpurity spectrum.

4. DISCUSSION

The above expression for the reciprocal relaxation time for phonons scattered by pair quadrupole orthocomplexes makes it possible to take into account the contribution of this mechanism to the thermal conductivity coefficient κ of solid hydrogen:

$$\kappa=\tfrac{1}{3}C_Vs^2\tau,\tag{39}$$

where C_V is the phonon-gas specific heat. The total effective mean free time of phonons, τ, is determined by the interference between the normal and U-processes and the scattering by the sample boundary, grain boundaries, and impurity centers in the crustal. Generally, for τ we have the Callaway formula[21]

$$\tau=\langle\tau_t\rangle+\frac{\langle\tau_t/\tau_N\rangle^2}{\langle1/\tau_N\rangle-\langle\tau_t/\tau_N^2\rangle},\tag{40}$$

in which

$$\frac{1}{\tau_t}=\frac{1}{\tau_N}+\frac{1}{\tau_U}+\frac{1}{\tau_B}+\frac{1}{\tau_p},$$

where τ_N and τ_U refer to N- and U-processes, respectively, and τ_B is the time of phonon relaxation due to scattering by boundaries. The angle brackets in (40) stand for averaging with the weighting function $w(q)$:

$$w(q)=\left[\frac{\omega(q)}{2T}\right]^2\left[\sinh\frac{\omega(q)}{2T}\right]^{-2},$$

$$\langle\cdots\rangle=\frac{1}{C_V}\int\frac{d^3q}{(2\pi)^3}(\cdots),\quad C_V=\int\frac{d^3q}{(2\pi)^3}w(q).$$

The orthopair concentration n_p in the expression for the relaxation time can be found from the conditions for equilibrium of the hydrogen orthosubsystem at a given temperature. Let us assume for definiteness that orthohydrogen is present in p-H_2 either in the form of single molecules (of concentration n_s) or in the form of pair clusters; we ignore the fact that the lattice may contain clusters consisting of large numbers of orthomolecules. Thus, the total average concentration of molecules of the ortho-modification in the crystal is $x=n_s+2n_p$. The value of n_p can be obtained from the following considerations. Let μ be the chemical potential of the system. Then at low concentrations x the condition that the free energy must be a minimum in μ yields

$$x=3e^{\mu/T}+z\,Z_0e^{2\mu/T},$$

where z is the coordination number. Here the first term on the right-hand side describes the concentration of single orthomolecules, with the factor 3 being the partition function of a single impurity that has a three-fold degenerate level $E=0$ (if the splitting due to the crystal field is ignored). The second term corresponds to the concentration $2n_p$ of the orthomolecules that form pair clusters. As a result we have

$$n_p=\frac{x}{2a}(\sqrt{1+a}-1)^2,\quad a=\frac{4}{9}xzZ_0. \tag{41}$$

The same expression for the pair-defect concentration can be obtained from the formulas of Refs. 10 and 11, with the only difference that a in (41) is set at $a=(4/9)Z_0P_p/P_s^2$, where $P_s=(1-x)^z$ is the probability of an orthomolecule being a single impurity, and $P_p=1-P_s$ is the probability of a molecule being a constituent of a pair defect. This expression for a reduces to that in (41) if we put $x\ll1$. Similar calculations can be carried out in the case where in addition to pairs the system has clusters consisting of large numbers of particles.

Now let us use (34) and (41) to calculate the temperature dependence of the average reciprocal relaxation time $\langle\tau_p^{-1}\rangle$ at pair orthocomplexes. This parameter is proportional to the total cross section of phonon scattering by a pair cluster and hence gives an unambiguous idea about the properties of the relaxation mechanism under investigation. The results of our numerical calculations are depicted by the solid curve *1* in Fig. 1. For the sake of comparison, in the same diagram we give the temperature curve of the average reciprocal relaxation time $\langle\tau_s^{-1}\rangle$ at single orthomolecules that has been constructed via the formulas of Ref. 9 (curve *2*), and the function $\langle\tau_p^{-1}\rangle$ reproduced from the data of Ref. 15 (curve *3*). All of the curves are calculated for the same value of the concentration of the orthocomponent, 5%. It can be seen that the quantity $\langle\tau_p^{-1}\rangle$ we calculated is two orders of magnitude smaller than the value $\langle\tau_s^{-1}\rangle$. A direct computer calculation of the thermal conductivity (Eq. (39)) of a solid ortho–para solution that uses formula (34) for the relaxation time and formula (41) for n_p shows that the contribution of orthopairs to the thermal conductivity of H_2 in the $T<5$ K range proves to be very small even in comparison to boundary scattering effects, while at higher temperatures it plays an insignificant role in comparison to the dominant contribution of *N*- and *U*-processes.

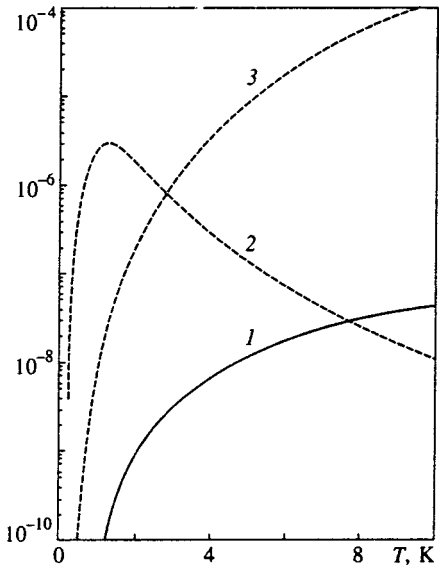


FIG. 1. Temperature curves for the average reciprocal relaxation time for the inelastic scattering of phonons by pair orthoclusters in solid hydrogen at an orthocomponent concentration $x=5$ %: curve *1* represents our results and curve *3*, Kokshenev's results.[15] Curve *2* represents the temperature behavior of the reciprocal time of relaxation at single orthomolecules calculated from Ebner and Sung's data.[9] All times are given in units of the reciprocal Debye frequency.

Generally speaking, this results has an obvious physical meaning. Indeed, a pair orthocluster in the zeroth approximation is simply a scattering orthocenter of double strength. It is clear from general considerations that the effects of excitation of the internal degrees of freedom, effects that arise because of the perturbation of the spectrum due to the interaction of the single molecules comprising the stable orthopair, constitute only a small correction to the double contribution of the single molecules at the neighboring sites of the matrix. Transitions in the orthocluster spectrum occur because the energy of the quadrupole–quadrupole interaction V_Q and the translational part of the phonon kinetic energy do not commute. At low temperatures, long-wavelength low-energy phonons dominate in the lattice, which in the final analysis leads to a small cross section of the inelastic scattering of the photons by orthopairs. For $T<T_0$ the phonon energy is insufficient for exciting resonant transitions, while for $T>T_0$ resonant scattering is possible, but in this range its contribution is suppressed by the already substantial contribution of *N*- and *U*-processes.

However, in this connection we must mention the remarkably large value of $\langle\tau_p^{-1}\rangle$ obtained by Kokshenev in Ref. 15, exceeding $\langle\tau_s^{-1}\rangle$ in the 1–5 K range by a factor of 100. No less remarkable is the temperature dependence of the reciprocal inelastic-scattering time: it has a peak near 1 K and rapidly decreases with increasing temperature, while it would seem that the probability of inelastic processes should increase. The explanation is that when calculating the deformation potential generated by phonons for the pair orthoclus-

ter, Kokshenev[15] allowed only for the part of the potential that is related to variations in the orientation of the vector **n**. What he overlooked was the fact that under rotation of the orthopair as a whole through an angle φ the vectors $\mathbf{S}_i$ also acquire increments, and this is true even if the orthomolecules do not change their orientation with respect to **n**. If the perturbation potential is now calculated with allowance of the discarded terms, it proves to be zero to within terms that are a reflection of the change in the bond length in the orthopair and contribute nothing to the inelastic scattering (as noted earlier). Since Kokshenev[15] allowed neither for the corrections to elastic scattering related to changes in R_0 nor for the term $\Delta H_{\rm kin}$ (see Eq. (6)), the approach he used must yield no inelastic scattering by pair clusters in the hydrogen paramatrix.

Thus, our study suggests that the low-temperature anomalies in the thermal conductivity of solid hydrogen observed in experiments probably cannot be explained solely by the contribution of the scattering of phonons by orthomolecules. At the same time it is obvious that the thermal conductivity of the paramatrix depends significantly on the concentration x of the orthomolecules present in the matrix. Hence the natural question is: Can orthoimpurities have an indirect effect on other relaxation mechanisms? In particular, Bohn and Mate[5] proposed a phenomenological model according to which the contribution of boundary scattering increases monotonically with x. But why then does an increase in orthoimpurity concentration lead to an increase in the surface area of the boundaries, i.e., to a decrease in the crystallite size in the sample? An alternative approach is to allow for the scattering by other structure defects, e.g., dislocations. To check whether this is the case we did a preliminary calculation of the contribution of dislocations to the low-temperature thermal conductivity of the H_2 matrix, using the well-known results of Klemens.[3] If we assume that the dislocation density ρ increases with the orthomolecule concentration, then we can give a satisfactory description of the low-temperature branch of $\kappa(T)$ as a function of x, and as x changes from 0.05 to 5%, ρ increases by a factor of approximately five, retaining a reasonable order of magnitude: $\rho \sim 10^8$–$10^9\ \mathrm{cm}^{-2}$. For comparison we note that in the same orthocomponent concentration interval the size of the crystallites predicted in Ref. 5 decreases by a factor of approximately 50. Of course, we still do not know the physical reasons for a relationship between ρ and x, the more so that the defect structure of samples used in measuring thermal conductivity is not monitored and the literature has no data on the density of dislocations in H_2. Nevertheless, using the numerous and well-documented data on the dislocation structure of classical HCP-crystals with impurities,[22,23] we can make reasonable assumptions concerning the increase in dislocation density and number of stacking faults in solid hydrogen with the orthocomponent concentration in the paramatrix. Naturally, these ideas are only qualitative, and establishing the role of lattice defects in thermal conductivity requires further studies.

In conclusion we note that our results are of a general nature and can be used to examine the kinetic properties not only of solid hydrogen but of other molecular cryocrystals as well. The real systems in which the mechanism discussed above may strongly influence the observed low-temperature thermal properties are, in particular, solid solutions of the molecules CO_2, CO, N_2, and N_2O in argon and krypton matrices.[4] All these impurities have large quadrupole electric moments, so that, e.g., the fraction of quadrupole–quadrupole interaction amounts to more than 60% of the total energy of the intermolecular interaction of carbon-dioxide impurities in a $Ar–CO_2$ solution, and it can be expected that the mechanism discussed in this paper provides a considerable contribution to the thermal conductivity of such mixtures. There is also another class of problems to which our results can be applied. The internal degrees of freedom of quadrupole clusters can be excited not only by phonons but also by fields of inner stresses generated by defects in the crystal structure, e.g., dislocations. Thus, inelastic relaxation at quadrupole defects must manifest itself not only in the thermal properties of solid cryosolutions but also in the mechanical properties. Allowing for this fact may be important in interpreting some experimentally observable effects, such as the thermomechanical effect in the thermal expansion of solid solutions of the $Ar–N_2$-type.[24]

*)E-mail: chishko@ilt.kharkov.ua

[1] J. M. Ziman, *Electrons and Phonons*, Oxford Univ. Press, Oxford (1960).
[2] J. Reissland, *The Physics of Phonons*, W. A. Benjamin, Reading, MA (1975).
[3] P. G. Klemens, in *Solid State Physics: Advances in Research and Applications*, H. Ehrenreich, F. Seitz, and D. Turnbull (eds.), Vol. 7, Academic Press, New York (1958), p. 1.
[4] *Physics of Cryocrystals*, V. G. Manzhelii and Yu. A. Freiman, AIP Press, New York (1997).
[5] R. G. Bohn and C. F. Mate, Phys. Rev. B **2**, 2121 (1970).
[6] J. E. Huebler and R. G. Bohn, Phys. Rev. B **17**, 1991 (1978).
[7] C. L. Reynolds Jr. and A. C. Anderson, Phys. Rev. B **22**, 5547 (1980).
[8] B. Ya. Gorodilov and V. B. Kokshenev, J. Low Temp. Phys. **81**, 45 (1990).
[9] C. Ebner and C. C. Sung, Phys. Rev. B **2**, 2115 (1970).
[10] J. F. Yarvis, H. Meyer, and D. Ramm, Phys. Rev. **178**, 1461 (1969).
[11] H. Meyer, Phys. Rev. **187**, 1173 (1969).
[12] T. Nakamura, Prog. Theor. Phys. **14**, 135 (1955).
[13] T. N. Antsygina, B. Ya. Gorodilov, N. N. Zholonko, A. I. Krivchikov, V. G. Manzheliĭ, and V. A. Slyusarev, Fiz. Nizk. Temp. **18**, 417 (1992) [Sov. J. Low Temp. Phys. **18**, 283 (1992)].
[14] T. N. Antsygina and V. A. Slyusarev, Fiz. Nizk. Temp. **19**, 494 (1993) [Low Temp. Phys. **19**, 348 (1993)].
[15] V. B. Kokshenev, J. Low Temp. Phys. **20**, 373 (1975).
[16] F. J. Dyson, Phys. Rev. **73**, 929 (1948).
[17] D. N. Zubarev, Usp. Fiz. Nauk **71**, 171 (1960).
[18] W. V. Lovitt, *Linear Integral Equations*, McGraw-Hill, New York (1924) (Reprint: Dover, New York (1950)).
[19] A. A. Maradudin, I. P. Ipatova, E. W. Montroll, and G. H. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation*, 2nd ed., Academic Press, New York (1971).
[20] V. F. Sears and J. van Kranendonk, Can. J. Phys. **42**, 980 (1964).
[21] J. Callaway, Phys. Rev. **113**, 1146 (1959).
[22] J. Friedel, *Dislocations*, Pergamon Press, Oxford (1964).
[23] J. P. Hirth and J. Loethe, *Theory of Dislocations*, McGraw-Hill, New York (1968).
[24] A. N. Aleksandrovskiĭ, K. A. Chishko, V. B. Esel'son, V. G. Manzheliĭ, and B. G. Udovidchenko, Fiz. Nizk. Temp. **23**, 999 (1997) [Sov. J. Low Temp. Phys. **23**, 750 (1997)].

Translated by Eugene Yankovsky

Exact, complete, and universal continuous-time worldline Monte Carlo approach to the statistics of discrete quantum systems

N. V. Prokof'ev,*) B. V. Svistunov, and I. S. Tupitsyn

Kurchatov Institute, 123182 Moscow, Russia
(Submitted 20 November 1997)
Zh. Éksp. Teor. Fiz. **114**, 570–590 (August 1998)

We show how the worldline quantum Monte Carlo procedure, which usually relies on an artificial time discretization, can be formulated directly in continuous time, rendering the scheme exact. For an arbitrary system with discrete Hilbert space, none of the configuration update procedures contain small parameters. We find that the most effective update strategy involves the motion of worldline discontinuities (both in space and time), i.e., the evaluation of the Green's function. Being based on local updates only, our method nevertheless allows one to work with the grand canonical ensemble and nonzero winding numbers, and to calculate any dynamical correlation function as easily as expectation values of, e.g., total energy. The principles found for the update in continuous time generalize to any continuous variables in the space of discrete virtual transitions, and in principle also make it possible to simulate continuous systems exactly. © *1998 American Institute of Physics.* [S1063-7761(98)01508-X]

1. INTRODUCTION

Quantum Monte Carlo (MC) simulation is the most powerful available method, if not the only one, of obtaining accurate results for complex systems, where analytic solutions are not possible and exact diagonalization methods do not work because of the enormous Hilbert space. However, most MC schemes are far from ideal, and suffer from significant shortcomings. These include (see, e.g., the most recent review article Ref. 1)

a) systematic errors due to artificial time discretization, which in most schemes scales as $(\Delta\tau)^2$, where $\Delta\tau$ is the time slice width;

b) restriction of the simulation to the zero winding number sector $M=0$ (a configuration in which world lines connect the initial state $|\alpha_1,\alpha_2,\ldots,\alpha_L\rangle$ at $\tau=0$ to the final state $|\gamma_1,\gamma_2,\ldots,\gamma_L\rangle$ at $\tau=\beta$, with the set $\{\gamma_i\}$ being obtained by cyclically permuting $\{\alpha_i\}$ M times (and all topologically equivalent configurations), is said to have a winding number M). Such a restriction results in systematic errors too, which however vanish with increasing system size. Also, one loses the ability to study topological excitations in the system, e.g., vortices or supercurrent states;

c) working with a fixed number of particles $N=\text{const}$ (canonical ensemble);

d) the critical slowing-down problem, which arises close to a second-order phase transition. This problem is closely related to constraints (b) and (c), and is indicative of inefficient procedures used to update configurations with large length scales;

e) slow accumulation of statistics when calculating correlation functions of operators not present in the initial Hamiltonian, e.g., the Green's function;

f) small acceptance rates in update procedures. These may be due to small parameters present in the formulation of the MC scheme, or systems described by Hamiltonians with different energy scales (e.g., when the hopping matrix element t is much smaller than the typical potential energy change $U\gg t$), or the necessity of global Metropolis updates, which arise in certain cluster-update algorithms;

g) anomalous dependence of the computation time on system size (due to self-averaging effects in the thermodynamic limit, the computation time required to achieve given accuracy is expected to be system-size independent);

h) a notorious sign problem, which emerges when the configuration weight is not positive definite. Since we do not see any reasonable solution of the sign problem in the general case, in what follows we exclude it from the discussion.

To eliminate some of these shortcomings, a number of different MC schemes were developed. Unfortunately, none of the existing schemes succeeded in solving all of them (leaving the sign problem aside) in the general case: there are extremely efficient algorithms which are far from universal, while the efficiency of existing universal algorithms is far from high for a large number of problems.

The standard worldline algorithm is based on imaginary time discretization and utilizes the small parameter $t\Delta\tau\ll 1$ in an approximate treatment of noncommuting operators in the Hamiltonian, known as Trotter break-up.[2,3] Physical intuitiveness and easy programming probably make this method the one most widely used. On the other hand, its weak points range over the whole list from (a) to (f), the most severe ones being (e) and (f).

In the worldline algorithm, one describes the configuration by specifying the system state $|\alpha_k\rangle$ at all time slices $\tau_k=k\Delta\tau$, where $k=0,1,\ldots,K_\beta$ and $\tau_{K_\beta}=1/T\equiv\beta$. The system state is then conventionally defined in the basis set in which the potential energy of the system is diagonal, i.e., in the site representation. Let us consider, as a typical example, the Hamiltonian of interacting particles on a lattice

$$H=-t\sum_{\langle ij\rangle} a_i^\dagger a_j+\sum_{ij} U_{ij}n_i n_j, \tag{1}$$

where $a_i^\dagger$ creates a particle at site i, t is the hopping matrix element, $n_i=a_i^\dagger a_i$, and $\langle ij\rangle$ denotes nearest-neighbor sites. From now on, we call points in time at which the system changes state ''kinks.'' The typical separation in time between two adjacent kinks at the same site is of order $1/t$ and independent of $\Delta\tau$, so that for small $\Delta\tau$ there are some $1/(t\Delta\tau)\gg 1$ time intervals between them.

The acceptance rate of the variation suggesting creation of a new kink–antikink pair is proportional to the square of the small parameter $(t\Delta\tau)^2$. On the other hand, when the MC procedure suggests shifting an already existing kink to the nearest point in time, the corresponding variation of the configuration is accepted with probability $\sim O(1)$. Thus, on average, by selecting different random time slices, it takes some $1/(t\Delta\tau)^2$ attempts to create a new kink–antikink pair and $1/(t\Delta\tau)$ attempts to move a kink to the nearest position in time. Still, the updated configuration is only slightly different from the previous one, and expectation values calculated before and after the variation are strongly correlated. An uncorrelated contribution of the given configuration fragment is obtained by shifting the kink a distance of order $1/t$, which requires some $1/(t\Delta\tau)^3$ operations, since the kink shift process is diffusive in nature. This means that the autocorrelation time in the standard worldline algorithm grows $\propto(\Delta\tau)^{-3}$ even in the absence of critical slowing down. Since all update procedures are local, the algorithm is subject to critical slowing down near the transition temperature.

In order to calculate the Green's function $\mathcal{G}(i,\tau)$, two worldline discontinuities are inserted at time slices $\tau_1=0$ and $\tau_2=\tau$ (in other words, one extra worldline is inserted in or removed from the interval $[\tau_1,\tau_2]$).[3,4] One then probes different configurations using standard update procedures and collects statistics in a d-dimensional histogram, which describes the spatial separation i between the discontinuities. The length of the time interval is then changed, and the same calculation is repeated. One MC step, i.e., the number of update operations performed between successive ''measurements,'' whereupon another point is included in the statistics of the calculated quantity, is proportional to $L^d\beta$, where L^d is the number of lattice sites considered. Thus, it requires about $L^d\beta$ operations to include only one point in the $(d+1)$-dimensional spacetime histogram for G.

In a sense, the standard worldline procedure of calculating $\mathcal{G}(i,\tau)$ has an anomalous dependence on N and β, since it takes at least $(N\beta)^2$ operations to update the whole histogram (typically G decays in space and time, and large scale behavior requires much more computation). We note that winding numbers and the grand canonical ensemble average can be incorporated, in principle, in the worldline algorithm. It is sufficient to consider separate contributions to the statistics of $\mathcal{G}(i,\tau)$ when $i=ML$ and $\tau=n\beta$ with integer M and n. However, in practice, only small systems at rather high temperatures can be considered using this algorithm.

The determinant method based on the Hubbard–Stratonovich transformation[5–7] also uses the discrete-time Trotter break-up, and thus becomes more and more inefficient due to long autocorrelation times when $\Delta\tau\to 0$ (points (a) and (f) above). It has an important advantage over the worldline method in calculating the Green's function, since it works with the grand canonical ensemble. However with increasing system size, the calculation time scales as L^3 (point (g)) and some of the procedures become ill-conditioned at low temperatures.

Another technique allowing Green's function calculations is called Green-function MC (or, more generally, the projection operator method).[8] It is applicable at zero temperature only, and the final result for $\mathcal{G}(i,\tau=0)$ depends on the trial wave function (we are not aware of whether it is possible to calculate the time dependence of $\mathcal{G}(i,\tau)$ by this method).

The stochastic series expansion (SSE) technique,[1,9,10] which stems from the Handscomb's method,[11] relies on the direct Taylor expansion of the statistical operator. This scheme is exact (contains no systematic errors). SSE has clearly demonstrated that time discretization is an artificial trick that is not at all necessary for MC simulation. Since an elementary update in the SSE scheme is equivalent to roughly $1/(t\Delta\tau)^3$ updates in the standard worldline method, it results in a significant drop in computation time for high-precision calculations. The rest of the problems, i.e., (b)–(f), survive in the SSE approach (point (f) still applies, because by expanding in powers of the full Hamiltonian one has to compare weights corresponding to the kinetic- and potential-energy terms, and if, e.g., $U\gg t$ in the Hamiltonian (1), then small acceptance rates appear in the update procedures). Still, away from the transition point, for large systems at low temperature, for which $U\sim t$, the SSE method is superior in evaluating basic thermodynamic properties like the total energy and density—density or current—current correlation functions.

A qualitatively new class of extremely efficient MC schemes[12] has been developed in recent years.[13–18] These schemes are based on the so-called loop cluster update (LCU) algorithm, which performs nonlocal updates for worldline loops with sizes as large as the system itself. Apart from solving the problem of critical slowing down, it also allows one to work in the grand canonical ensemble and with nonzero winding numbers. From this method we learn that problems (b), (c) and (d) can be circumvented. Unfortunately, the LCU algorithm, as far as we know, is not universal. It applies to spin systems and to hard-core Hubbard models, but was never formulated for the general lattice Hamiltonian, like interacting soft-core bosons with arbitrary U_{ij}, arbitrary density (chemical potential), or on-site disorder. Another shortcoming of the LCU, which in a sense can also be called nonuniversality, is that it does not admit of a universal code. It should be noted that LCU allows for considerable generalization to the cases of external magnetic field and disorder, but generally speaking the cost is lost efficiency because of the exponentially small acceptance rates for large loops[1)].

It is shown in Ref. 19 how to build a path integral in continuous time for quantum systems in a discrete basis. The configuration is specified by transition times and system states before and after the transition. Within this description

one can formally think about taking the limit $\Delta\tau\to 0$ in the standard approach. However, to implement this description one has to formulate the update process. In the standard worldline algorithm, one of the basic procedures is generation of new kink–antikink pairs when system evolution on a given site changes from $|\alpha\rangle\to|\alpha\rangle\to|\alpha\rangle$ to $|\alpha\rangle\to|\gamma\neq\alpha\rangle\to|\alpha\rangle$. In the continuous limit, the acceptance rate for such a variation vanishes as $(\Delta\tau)^2$, and thus the problem of qualitatively new update principles arises.

Recently, two independent continuous time schemes utilizing the ideas of Ref. 19 were developed.[18,20] Beard and Wiese[18] find that with the LCU algorithm, one can go directly to the continuous-time limit $\Delta\tau\to 0$, thus rendering the LCU algorithm *exact*.

The general solution to the problem of configuration update in continuous time is found in Ref. 20. The resulting continuous-time worldline (CTWL) method is exact, and like SSE, is $1/(t\Delta\tau)^3$ times more efficient than finite-$\Delta\tau$ local schemes. It also completely eliminates problem (f), since none of the procedures relies on small parameters (potential energy is accounted for in the exponent, and one does not have to weight relative contributions to the statistics of the potential and kinetic energy terms, as happens in SSE).

In its original formulation, the CTWL approach did not solve problems (b)–(e), and was tested only on a simple single-particle Hamiltonian.[20] In the present article, we present a complete description of the CTWL approach to the statistics of arbitrary many-particle system with discrete Hilbert space. We demonstrate that it enables one to solve problem (e) in a physically intuitive way by formulating the *local* update procedures in terms of the motion of two worldline discontinuities (in what follows we call them ''worms'') in space and time, i.e., in terms of a calculation of the Green's function. During one MC step (consisting of $N\beta/t$ operations) the whole histogram for $\mathcal{G}(i,\tau)$ is updated, which means that G is calculated as efficiently as, say, the total energy, and is not affected by point (g). Since $\mathcal{G}(i=[M_xL_x,M_yL_y,\ldots],n\beta)$ with integer M_x, $M_y,\ldots$ and n describes a system with n extra particles and winding numbers $\{M\}$, we are working in the grand canonical ensemble. This solves problems (b) and (c).

Closer examination of the loop building rules[14,18] for the Heisenberg Hamiltonian shows their remarkable similarity to the evolution of an extra worldline segment. The crucial difference is that only closed loops are considered by the LCU algorithm, while our scheme considers all the intermediate configurations as well, and utilizes them for the Green's function calculation. Working in the extended configuration space, which includes discontinuous worldlines, we use local Metropolis[21] updates only. However, when discontinuities annihilate, and we return to the configuration space of closed worldlines, the net result of the update is of global character. Since CTWL with ''worm'' updates effectively mimics single-loop LCU, we may hope that it possesses all the remarkable features inherent in LCU, and in particular that it solves, or at least softens, the problem (*d*).

To summarize, we propose a method which is exact, complete (allows calculation of any correlation function), and universal (applies to arbitrary quantum systems with discrete Hilbert space, and enables one to write a unified code, that is simultaneously applicable to lattice bosons and arbitrary spins, with arbitrarily long-range interactions and disorder). The sign problem now becomes the only stumbling block to making quantum MC an ''ideal'' computational tool for studying complex systems.

This paper is organized as follows. In Sec. 2 we formulate the general principles of the continuous-time worldline approach. In Sec. 3 we introduce the update procedures that we find to be the most effective and sufficient for simulation of the quantum statistics of many-particle systems. In Sec. 4 we demonstrate the advantages of the new method by presenting some results that cannot be obtained by any other MC approach: the Green's function and the critical index of the 1-*D* boson Hubbard model at the quantum critical point, and the low-energy properties of the strongly disordered Bose glass phase. In Sec. 5 (and Appendix A) we discuss the feasibility of increasing the efficiency of our method in the case of long-range interactions, and consider the feasibility of generalizing the method to continuous systems.

2. GENERAL PRINCIPLES

Let H_0 and V be the diagonal and off-diagonal parts of the Hamiltonian H in a chosen representation corresponding to the full set $\{\alpha\}$ of eigenstates of H_0, with $H_0|\alpha\rangle=E_\alpha|\alpha\rangle$. The statistical operator can then ordinarily be related to the Matsubara evolution operator σ in the interaction picture, i.e., we write $e^{-\beta H}=e^{-\beta H_0}\sigma$, with

$$\sigma=1-\int_0^\beta d\tau\, V(\tau)+\ldots +(-1)^m\int_0^\beta d\tau_m\cdots\int_0^{\tau_2}d\tau_1 V(\tau_m)\cdots V(\tau_1)+\ldots, \quad (2)$$

where $V(\tau)=e^{\tau H_0}Ve^{-\tau H_0}$. Without loss of generality and in accordance with typical forms of Hamiltonians of interest, V can be written as a sum of elementary terms Q_s, whose action on any function from the set $\{\alpha\}$ results in another function from this set:

$$V=\sum_s Q_s,\quad Q_s|\alpha\rangle=-q_{\gamma\alpha}(s)|\gamma\rangle\quad(\gamma=\gamma(s,\alpha)). \quad (3)$$

Since V is Hermitian, for any s in the sum (3) there exists an s' such that $Q_{s'}=Q_s^\dagger$. We rewrite Eq. (2) in components (below $E_{\alpha\gamma}\equiv E_\alpha-E_\gamma$):

$$\sigma_{\alpha\gamma}=\delta_{\alpha\gamma}+\sum_s\int_0^\beta d\tau\, q_{\alpha\gamma}(s)e^{\tau E_{\alpha\gamma}}+\ldots +\sum_{s_1,\ldots,s_m}\int_0^\beta d\tau_m\cdots\int_0^{\tau_2}d\tau_1\, q_{\alpha\nu}(s_m)\times e^{\tau_m E_{\alpha\nu}}\ldots q_{\lambda\gamma}(s_1)e^{\tau_1 E_{\lambda\gamma}}+\ldots\ . \quad (4)$$

Note that there is no additional summation over the indices of the intermediate complete sets (labeled by Greek letters), since these are defined in a unique way by configurations of $(s_1,s_2,\ldots,s_m)$.

We confine ourselves to the case of finite-range interaction, which is defined by the requirement that for each term s_1 of elementary operators $\{Q_s\}$ there exists only a finite number of terms s_2 for which the condition

$$[Q_{s_1}(\tau_1), Q_{s_2}(\tau_2)] = 0 \tag{5}$$

is not met. In finite-range interactions, the structure of the series (4) is drastically simplified, the simplification being of crucial importance for practical realization of our algorithm. From (5) it follows that up to an irrelevant change in the indexing of energies and matrix elements, one can ignore the chronological order of $Q_{s_1}(\tau_1)$ and $Q_{s_2}(\tau_2)$ in the evolution operator.

This suggests representing a general term of the series (4) in the following form. First, we introduce the notion of a ''kink of type s,'' which is characterized by a time τ, a matrix element $q_{\alpha\gamma}(s)$, and a diagonal energy difference $E_{\alpha\gamma}$. The former two we refer to as parameters of the kink. It is essential that (i) to obtain parameters of a kink one need not know explicitly the whole state $|\alpha\rangle$, or $|\gamma\rangle$—local information is enough; (ii) to specify a particular structure of a term in Eq. (4), including the chronological order of all noncommuting operators, it suffices to specify associated with each kink the neighbors, i.e., the noncommuting kinks nearest in time.

Now our goal is to describe in general terms a stochastic process that directly evaluates Eq. (4). For simplicity, we assume that all $q_{\alpha\beta}(s)$ are positive real numbers. (In many particular alternative cases, a straightforward generalization is possible, but usually at the expense of convergence.) Summations and integrations in Eq. (4) then can be regarded, up to a normalizing factor, as an averaging over the statistics of different configurations of kinks, each configuration being defined by a certain number of kinks of certain types, their associations and particular positions in imaginary time. The Monte Carlo process should examine these statistics by generating different kink configurations in accordance with their weights. The global process will consist of a number of elementary subprocesses, each being responsible for certain modifications of a particular type.

An update procedure of a general type should involve subprocesses of creation and annihilation of kinks. Clearly, the qualitative difference between discrete- and continuous-time QMC schemes is associated with processes of just this kind. To introduce the general principles of construction of subprocesses that change the total number of kinks, we consider some particular (but still rather rich) class of elementary transformations (which seems to be sufficient for all practical purposes). By an elementary transformation we mean a subprocess that either only creates or only annihilates a certain number of kinks. The set of elementary subprocesses can be decomposed into self-balanced creation–annihilation pairs. Our task then is to specify the structure of creation and annihilation subprocesses, and to derive the balance equation that would guarantee that the statistics generated by each pair of subprocesses does really correspond to that introduced by Eq. (4).

Let some subprocess create n kinks of given type $s_1, s_2, \ldots, s_n$, the temporal positions of the kinks being specified by the n-dimensional vector $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_n\}$. In the most general case, the creation procedure involves two steps.

First, one considers creating n new kinks at $\boldsymbol{\tau} \in \Gamma$, where Γ is a certain region in the n-dimensional space of times $\tau_1, \tau_2, \ldots, \tau_n$. The probability density $W(\boldsymbol{\tau})$ of choosing a given $\boldsymbol{\tau}$ is, generally speaking, arbitrary, provided $W(\boldsymbol{\tau})$ is nonzero at every physically meaningful configuration of kinks.

In the second step, one either accepts (with probability $P_{\rm acc}(\boldsymbol{\tau})$) or rejects the suggested modification.

The annihilation procedure is much simpler. The n kinks of given type $s_1, s_2, \ldots, s_n$ and with $\boldsymbol{\tau} \in \Gamma$ are either removed (with probability $P_{\rm rem}(\boldsymbol{\tau})$) or remain untouched.

The equation of balance for the given pair of subprocesses reads

$$A_0\, p_c W(\boldsymbol{\tau}) P_{\rm acc}(\boldsymbol{\tau}) d\boldsymbol{\tau} - dA_n(\boldsymbol{\tau}) p_a P_{\rm rem}(\boldsymbol{\tau}) = 0. \tag{6}$$

Here A_0 ($A_n(\boldsymbol{\tau})$) is the probability (probability density) of finding a configuration without the specified n kinks (with the specified n kinks at the given $\boldsymbol{\tau}$). We have also introduced the probabilities p_c and p_a of addressing the creation and annihilation subprocedures. In the next section we show how it can turn out quite naturally that these probabilities do not coincide.

The statistical interpretation of Eq. (4) implies

$$\frac{dA_n(\boldsymbol{\tau})}{A_0} = d\boldsymbol{\tau} \prod_{j=1}^{n} q(s_j) \exp(\Delta E_j \tau_j), \tag{7}$$

where $q(s_j) \equiv q_{\alpha_j\beta_j}(s_j)$ and $\Delta E_j \equiv E_{\alpha_j} - E_{\beta_j}$. Combining (6) and (7) we obtain the necessary and sufficient condition for the pair of subprocesses to be self-balanced:

$$\frac{W(\boldsymbol{\tau}) P_{acc}(\boldsymbol{\tau})}{P_{rem}(\boldsymbol{\tau})} = R(\boldsymbol{\tau}),$$

$$R(\boldsymbol{\tau}) = \frac{p_a}{p_c} \prod_{j=1}^{n} q(s_j) \exp(\Delta E_j \tau_j). \tag{8}$$

Given $W(\boldsymbol{\tau})$, the condition (8) is satisfied, e.g., by the following obvious choice of P_{acc} and P_{rem}.

$$P_{acc}(\boldsymbol{\tau}) = \begin{cases} R(\boldsymbol{\tau})/W(\boldsymbol{\tau}), & \text{if } R(\boldsymbol{\tau}) < W(\boldsymbol{\tau}) \\ 1, & \text{otherwise} \end{cases}, \tag{9}$$

$$P_{rem}(\boldsymbol{\tau}) = \begin{cases} W(\boldsymbol{\tau})/R(\boldsymbol{\tau}), & \text{if } R(\boldsymbol{\tau}) > W(\boldsymbol{\tau}) \\ 1, & \text{otherwise} \end{cases}. \tag{10}$$

From (9) it can be seen that there is a certain reason for choosing $W(\boldsymbol{\tau}) \propto R(\boldsymbol{\tau})$, as in this case P_{acc} becomes independent of τ, and the accept—reject decision can be made before suggesting a particular configuration, thus saving computational time. However, if the structure of the function $R(\boldsymbol{\tau})$ is complicated, the numerical generation of the corresponding distribution will be very expensive. In this case it is better to take $W(\boldsymbol{\tau}) \propto \tilde{R}(\boldsymbol{\tau})$, where $\tilde{R}(\boldsymbol{\tau})$ is some ''coarse-grained'' approximation to $R(\boldsymbol{\tau})$ with a simple form.

We do not consider here a general theory of subprocesses that do not change the number of kinks, since it is basically the well-known theory of taking multidimensional

integrals by standard Monte Carlo procedures. Particular examples of such subprocesses can be found in Ref. 20 and in the next section.

The foregoing approach does not involve any explicit truncation of the series (4). One might wonder, however, what the effect of implicit truncation in the practical realization of the process would be, due to the finite size of the computer memory. To this end we note that even for simulations of many-particle systems, where the typical number of kinks $N_{\rm kink}$ (that is, the typical number of terms in the series (4)) that contribute to the final result is really large, and one might expect the memory/accuracy problem, the effect can be easily made absolutely negligible. Indeed, from the Central Limit Theorem, it follows that the number of kinks in significant configurations has a Gaussian distribution with the peak at $\bar{N}_{\rm kink}$ and a half-width of order $\sqrt{N_{\rm kink}}$ (cf. Ref. 9). If one just reserves at least twice as much memory as necessary to describe the configuration with $\bar{N}_{\rm kink}$ elements, then during a computation spanning the age of the Universe, the system will not fluctuate to states which cannot be fit into memory. The implicit truncation error thus can be made astronomically small.

3. UPDATE PROCEDURES

A. Kink motion

Let us first consider update procedures that are straightforward generalizations of those known in the discrete-time worldline algorithm, and work with closed trajectories only. The simplest process involves transformations that do not change the number of kinks, but change their types, time positions, and temporal ordering,[20]

$$\langle\alpha|Q_{a_1}(\tau_{a_1})Q_{a_2}(\tau_{a_2})\ldots Q_{a_n}(\tau_{a_n})|\gamma\rangle \rightarrow\langle\alpha|Q_{b_1}(\tau_{b_1})Q_{b_2}(\tau_{b_2})\ldots Q_{b_n}(\tau_{b_n})|\gamma\rangle. \tag{11}$$

The number of operators involved in the transformation, their types, and time positions are not constrained, except that the two configurations have nonzero weight. Obviously, one could suggest many different realizations of Eq. (11), and some might work more efficiently than others, depending on the system. Here we describe the procedure called ''kink motion;'' other procedures have too much in common to be described separately, and allow trivial modifications.

To move a kink, we first select it at random from the list of existing kinks and decide on the time interval to be considered. Suppose that we have chosen a transition described by (Q_0,τ_0). We then find kinks of the same type that are nearest in time (both to the left and to the right of τ_0), i.e., Q_0 or $Q_0^\dagger$, and consider their times $\tau_1<\tau_0$ and $\tau_2>\tau_0$ as the boundaries of the time ''window'' transformed by this procedure (in certain configurations at high temperature, it may happen that $(\tau_1,\tau_2)=(0,\beta)$). It is allowed to have any number of kinks of different types $Q_a\neq Q_0,Q_0^\dagger$ within (τ_1,τ_2). Thus the typical initial configuration has the form

$$\ldots|_{\tau_1}Q_{a_1}(\tau_{a_1})Q_{a_2}(\tau_{a_2})\ldots Q_0(\tau_0)\ldots Q_{a_n}(\tau_{a_n})|_{\tau_2}\ldots, \tag{12}$$

(as explained above, one has to consider only those kinks which do not commute with Q_0).

The second step is to analyze all possible configurations obtained from (12) by removing Q_0 from point τ_0 and inserting it at arbitrary $\tau'\in(\tau_1,\tau_2)$. We keep the time positions and the chronological ordering of all the other operators $Q_{a_1},Q_{a_2},\ldots,Qa_n$ untouched. The new position of the selected kink Q_0 in time is decided according to the statistical weight of the final configuration as defined by Eq. (4). This is done in complete analogy with the classical MC procedure of taking multidimensional integrals.

The acceptance rate of the kink motion procedure is unity, since the differential measure of the initial configuration is zero. In this way, all noncommuting kinks in the Hamiltonian (except kink–antikink pairs, which are dealt with in the next subsection) can change places. In dimensions $d>1$, the kink motion procedure must be supplemented with a ''local loop'' procedure, which generates small loops in real space, e.g., by replacing $Q_{i\rightarrow i+g_1}(\tau_1)Q_{i+g_1\rightarrow i+g_1+g_2}(\tau_2)\rightarrow Q_{i\rightarrow i+g_2}(\tau_3)Q_{i+g_2\rightarrow i+g_1+g_2}(\tau_4)$, where g_1,g_2 are the nearest neighbor indices.

B. Creation and annihilation of kink–antikink pairs

In this subsection we make use of the general theory of Sec. 2 and explain how the elementary procedure of creation and annihilation of kink–antikink pairs is organized in practice. An important new principle realized in our algorithm is the possibility of selecting different update procedures with certain probabilities (see also Appendix A). These probabilities, p_a and p_c, are at our disposal, and if necessary, can be used to ''fine tune'' the efficiency of the MC process as a whole. The most natural starting point for the update is to address at random some configuration fragment. It can be characterized by the kink $Q_0(\tau_0)$, or by the system state $|\alpha(i_0)\rangle$ between the two adjacent kinks that change this state (in computer memory, all $|\alpha(i_0)\rangle$ between kinks are assigned labels; the configuration itself is described as a linked graph by specifying nearest-neighbor associations (in space and time) between the labels). We choose the latter variant and address site labels. Thus the probability of applying an update procedure to a given fragment is $\propto 1/N_{\rm lab}$ where $N_{\rm lab}$ is the total number of labels characterizing the initial configuration. By inserting (deleting) n extra kinks, we increase (decrease) $N_{\rm lab}$ by

$$\sum_{j=1}^{n} m_{Q_j},$$

where m_{Q_j} gives the number of states changed by the kink Q_j. Thus, the ratio p_a/p_c in Eq. (8) is proportional to

$$N_{\rm lab}\Big/\left(N_{\rm lab}+\sum_{j=1}^{n} m_{Q_j}\right),$$

when addressing the creation of n kinks, and

$$\left(N_{\rm lab}-\sum_{j=1}^{n} m_{Q_j}\right)\Big/ N_{\rm lab},$$

when addressing the annihilation procedure.

To fix the values of p_a and p_c, we count the number of possible kink–antikink processes that can be applied to a given fragment. This number is denoted by $N_{\rm proc}$. The simplest choice is then to assign equal weight, $1/N_{\rm proc}$ to all possibilities. For example, if we consider a model with the nearest-neighbor hopping in $1D$, then there are three possibilities for the site state $|\alpha(i)\rangle$: to insert $Q_{i\to i+1}Q_{i+1\to i}$ or $Q_{i\to i-1}Q_{i-1\to i}$ and to delete a pair of kinks that change this state to the left and to the right in time, provided they form a kink–antikink pair (i.e., are of the $Q_{i\pm1\to i}Q_{i\to i\pm1}$ type). In this case $m_{Q_j}=2$ as well, and we finally have

$$\frac{p_a}{p_c}=\frac{N_{\rm lab}}{N_{\rm lab}+4}\quad\text{(creation)};$$

$$\frac{p_a}{p_c}=\frac{N_{\rm lab}-4}{N_{\rm lab}}\quad\text{(annihilation)}. \tag{13}$$

Obviously, in the thermodynamic limit and at low temperature, these ratios are very close to unity. Again, this is only a particular example; other choices may prove to be more efficient under certain conditions.

Once the configuration fragment and update procedure are selected, we proceed along the lines described in Sec. 2. Here we would like to comment on the choice of probability density $W(\boldsymbol{\tau})$. It would be perfect from the acceptance rate point of view to take $W(\boldsymbol{\tau})\propto R(\boldsymbol{\tau})$. However, this can turn out to be a very expensive procedure. To illustrate the point, consider a configuration fragment of length $\tau_{l,r}=\tau_r-\tau_l$. Due to the large interaction radius between particles, an effective field acting on updated states can change many times during $\tau_{l,r}$. If the number of time slices thus induced on the interval (τ_l,τ_r) is $N_{\tau_{l,r}}\gg1$, then complete parametrization of the $R(\boldsymbol{\tau})$ function will require calculation of the $N_{\tau_{l,r}}(N_{\tau_{l,r}}+1)/2$ partial probabilities, according to the number of ways one can distribute two kinks among $N_{\tau_{l,r}}$ time subintervals.

The solution of the problem lies in choosing $W(\boldsymbol{\tau})=W(\bar{E},\tau)$, where $W(\bar{E},\boldsymbol{\tau})$ is an analytic function with the same properties as $W(\boldsymbol{\tau})$, controlled by a parameter $\bar{E}$ that is used to minimize the variance of $|W(\bar{E},\boldsymbol{\tau})-R(\boldsymbol{\tau})|$. The most obvious physical choice of $\bar{E}$ is the mean field potential acting on the updated states during $\tau_{l,r}$ from the rest of the system

$$\exp(-\bar{E}\tau_{l,r})=R(\tau_l,\tau_r), \tag{14}$$

$$W(\bar{E},\boldsymbol{\tau})=\frac{\exp(-\bar{E}(\tau_2-\tau_1))}{I},$$

$$I=\int_{\tau_l}^{\tau_r}d\tau_2\int_{\tau_l}^{\tau_2}d\tau_1\,\exp(-\bar{E}(\tau_2-\tau_1)). \tag{15}$$

One immediately recognizes in $W(\bar{E},\boldsymbol{\tau})$ the statistics of the kink–antikink pair in the biased two-level system,[20] which, through the mean-field definition of the bias energy $\bar{E}$, most closely approximates the local statistics of kink–antikink pairs in a real system.

The procedures described in the last two subsections represent a direct generalization of local procedures already known in the discrete-time worldline method. Their continuous-time versions are, however, only specific realizations of a much wider class of possible, procedures, thus making the overall CTWL scheme more flexible.

C. Creation—annihilation, jump, and reconnection procedures for worldline discontinuities

Up to now, we have considered procedures for working with closed worldlines. These are sufficient to simulate quantum statistics in the canonical ensemble and in the $M=0$ sector. To overcome this essential drawback, and to calculate the Green's function, one usually introduces an extra worldline segment and simulates quantum statistics in the presence of two worldline discontinuities at points (i_1,τ_1) and (i_2,τ_2). This process is highly inefficient, because one has to probe all degrees of freedom in the configuration (numbering roughly $\sim L^d\beta$) to collect statistics for only two extra degrees of freedom. In practice, this method was never used to calculate Green's function in large systems, e.g., with $L^d\beta\sim10^4$. The solution we find for this problem is in considering the two worldline discontinuities to be real dynamic variables in the Hamiltonian, which are allowed to move through the configuration both in space and time. It turns out that this motion can be arranged to be ergodic, and probes all possible system states. One can even completely ignore all the other update procedures, such as moving other kinks and working with kink–antikink pairs, probably at the expense of being less efficient, but still remaining accurate, complete, and universal. Below we describe the details of update procedures with worldline discontinuities (''worms''), which were first introduced in Ref. 20.

We start with the general expression for the Matsubara Green's function (see, e.g., Ref. 22) in the interaction picture

$$\mathcal{G}(i,j,\tau_1,\tau_2)=-e^{\beta\Omega}\,\mathrm{Tr}[e^{-\beta H_0}T_\tau(a_i(\tau_1)a_j^\dagger(\tau_2)\sigma)], \tag{16}$$

where T_τ is the τ-ordering operator, which was explicitly written before in defining the Matsubara evolution operator σ in Eq. (2); Ω is the grand canonical potential. To be specific, we assume here that H_0 is diagonal in the site representation; in the general case, one might imagine that the index i refers to some parametrization of eigenstates of H_0. Since we now work in the grand canonical ensemble, the Hamiltonian contains an extra term

$$-\mu N\equiv-\mu\sum_i n_i, \tag{17}$$

where μ is the chemical potential. Formally, the only difference between the statistics given by Eq. (4) and the Green's function (16) is that we have two extra kinks, $a_i(\tau_1)$ and $a_j^\dagger(\tau_2)$. Hence one has the possibility of calculating the Green's function in a *unified process*, together with standard thermodynamic averages (''energy,'' for the sake of brevity). To this end, it is necessary just to work in an extended configuration space, where two classes of configurations are present: (i) with continuous worldlines, and (ii) with two worldline discontinuities, corresponding to the kinks $a_i(\tau_1)$ and $a_j^\dagger(\tau_2)$. (Clearly, configurations of class (i) contribute to ''energy,'' while those of the class (ii) contribute to the

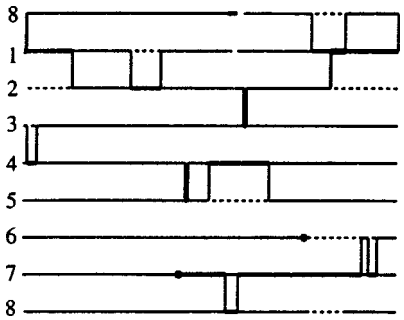


FIG. 1. A typical 8-site configuration with two worldline discontinuities marked by filled circles. The width of the solid line is proportional to the site occupation number, and dashed lines are empty sites.

Green's function.) The transitions between the two classes are performed by the processes of creation and annihilation of the kinks $a_i(\tau_1)$ and $a_j^\dagger(\tau_2)$, in accordance with the general balance principles Eq. (8). For computational purposes, it is reasonable to redefine the Green's function by a trivial scaling transformation $a_i \to \eta^* a_i$, $a_j^\dagger \to \eta a_j^\dagger$, where the constant η is adjusted to produce the optimal acceptance (rejection) probability.

Alternatively, one can arrive at the above scheme by the standard trick of introducing a source to the configuration action S (the notation η for the source is chosen deliberately):

$$\int_0^\beta d\tau V(\tau) \to \int_0^\beta d\tau V(\tau) + \sum_i \int_0^\beta d\tau \times (\eta_i^*(\tau) a_i(\tau) + \eta_i(\tau) a_i^\dagger(\tau)), \tag{18}$$

and defining the Green's function as a functional derivative of the generating functional (the partition function with the source)

$$\mathcal{G}(i,j,\tau_1,\tau_2) = -\frac{1}{Z} \left. \frac{\delta^2 Z}{\delta\eta_i(\tau_1)\delta\eta_j^*(\tau_2)} \right|_{\eta,\eta^*\to 0}. \tag{19}$$

The numerical procedure equivalent to the variational derivative in the limit $\eta \to 0$ means that only configurations with (i) zero and (ii) two worldline discontinuities are included in the statistics. Confining ourselves to just these configurations, we do not have to deal any longer with infinitesimally small η, and can choose η to be a certain finite constant. (This is crucial for any realistic computational process, since $\eta \to 0$ clearly means that the time of accumulation of statistics goes to infinity.) Indeed, a particular value of η just defines the relative weights of classes (i) and (ii), thus changing the relative norm of the Green's function with respect to ''energy'' by the known factor of $|\eta|^2$. (Incidentally, one may pay no attention at all to the normalizing statistics for the Green's function, as the norm can ultimately be fixed by the condition $\mathcal{G}(i,i,\tau,\tau+0) = -\text{density}$.)

A typical configuration with two 'worms'' is shown in Fig. 1 (''live'' picture taken from the computer). To update it we apply the following transformations:

Creation and annihilation of two worldline discontinuities. We delete a pair, $a_i(\tau_1)a_i^\dagger(\tau_2)$ or $a_i^\dagger(\tau_1)a_i(\tau_2)$, when discontinuities happen to meet at the same site i and there are no other kinks between them that can change the state of i. The only difference between this and the kink–antikink procedure is that now we transform only a single-site state, thus $m_Q = 1$. The annihilation procedure addresses the pair of worms, but the creation procedure (which makes sense only when there are no worms) addresses the randomly selected configuration fragment label. The ratio of probabilities p_a/p_c to address update procedures that transform the same configuration fragment back and forth is now

$$\frac{p_a}{p_c} = N_{lab} \ \text{(creation)}, \quad \frac{p_a}{p_c} = N_{\text{lab}} - 2 \ \text{(annihilation)}. \tag{20}$$

This ratio is macroscopically large, which is obviously unpleasant for the computational process. However, we have $R(\tau) \sim |\eta|^2$, with the freedom of choosing η. By setting $|\eta|^2 \sim 1/\langle N_{\text{lab}} \rangle$, where $\langle N_{\text{lab}} \rangle$ is the average number of labels in the configuration, we obtain an update procedure that is not based on small parameters (in practice, any rough estimate like $(L^d\beta)$ for $\langle N_{\text{lab}} \rangle$ is sufficient). The rest is done in exactly the same manner as described in Sec. 3 B.

Jump. This update procedure is illustrated in Fig. 2. We select one of the worldline discontinuities and suggest shifting it in space by inserting an ordinary kink (hopping operator) to the left (in time) of the annihilation operator and to the right of the creation operator. As a result, the worm ''jumps'' to another site. The number of kinks changes by one in this procedure, but p_a/p_c is unity, because we address it upon the availability of worms, and not according to the number of labels. Also, since we are dealing with only one extra kink here, the structure of the $R(\tau)$ function (see Sec. 2) is much simpler, and we choose $W(\tau) = R(\tau)/\int d\tau R(\tau)$. The integral is over the time interval of the updated fragment. The opposite procedure is called an ''anti–jump.''

Reconnection. Formally, this update procedure, which is shown in Fig. 3, is technically identical to the ''jump,'' but now an extra kink is inserted to the right of the annihilation operator and to the left of the creation operator. We still distinguish between them, because in the jump procedure the corresponding particle trajectories do not exchange places, while they do so in the reconnection update. Figure 3 makes it clear that we have effectively reconnected worldline seg-

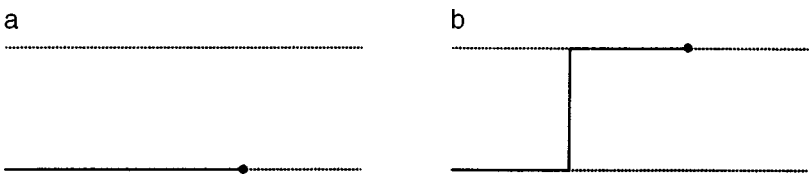


FIG. 2. Jump procedure for the annihilation operator. a) Initial configuration fragment; b) suggested variation (in the antijump procedure, (b) is the initial configuration and (a) is the suggested variation).

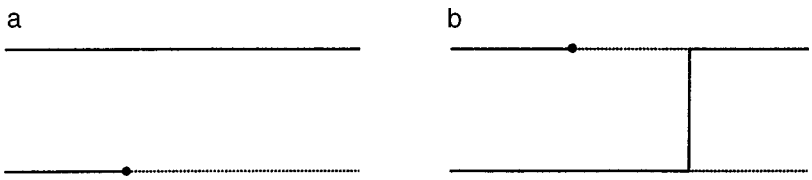


FIG. 3. Reconnection procedure for the annihilation operator. a) Initial configuration fragment; b) suggested variation (in the antireconnection procedure, (b) is the initial configuration and (a) is the suggested variation).

ments of different trajectories. Note that in fermionic systems, any reconnection/antireconnection procedure results in a change of the configuration sign.

Shift in time. The motion of worldline discontinuities in time is essentially the kink motion process (see Sec. 3 A). Suppose that we have decided to shift an annihilation kink $Q_0 = a_i$. The only difference from the scheme (12) is in the definition of the updated time interval. Its boundaries (τ_1, τ_2) now correspond to the time positions of the nearest left and right neighbors (kinks) of any type that operate on the same state *i*. Of course other possibilities are allowed as well, if one has some physical arguments in favor of, say, extending the time window farther to the next-nearest kink, or a kink of a special type.

The update procedures thus defined comprise an ergodic stochastic process that operates on the entire configuration space of the system. All configurations, including those with nonzero winding numbers and different number of particles, are accounted for. Extra particles are inserted/removed from the system when $a\dagger_j(\tau_2)$ makes a complete loop in time (relative to $a_i(\tau_1)$), i.e., when $\tau_2 - \tau_1$ changes by multiples of β. Winding numbers are introduced when $j - i$ changes by multiples of L. The key point of our approach is that each local update makes a contribution to the $\mathcal{G}(i,\tau)$ histogram, except rare cases in which there are no worms in the configuration; these configurations contribute to the ''diagonal'' (or conventional) statistics of closed worldlines. Contrary to the standard calculation, we do not adjust all degrees of freedom to the current positions of worms, but rather probe and update the whole configuration through their motion. This almost trivial modification results in a factor of $(L^d\beta)$ acceleration of the scheme!

It is instructive to draw an analogy between the motion of worldline discontinuities and the loop cluster update rules. As is easily seen, the basic elements of the single-loop LCU method known as ''optional decay'' and ''forced transition'' [18] correspond to a particular evolution of the worldline discontinuities (''optional decay'' corresponds to the ''jump'' procedure, and ''forced continuation'' to the ''antireconnection'' procedure). A closed loop is obtained after annihilating the pair a and $a\dagger$. Notice, however, that in our scheme (i) not only closed loops, but also all intermediate configurations are physically meaningful and are included into the statistics; (ii) nothing is based on the special structure of the system Hamiltonian; (iii) the update is always local (it is known that acceptance rates for large loops become very small when an external magnetic field in the *Z*-direction is applied to the Heisenberg system (magnetic field is equivalent to a finite chemical potential in bosonic language); this problem is simply absent in local schemes).

The statistics of discontinuities in space and time is given by $\mathcal{G}(i,\tau)$, i.e., it is defined by the Hamiltonian. In general, the optimal update scheme depends on the quantity being calculated, and thus one might wish to control the statistics of worldline discontinuities ''at will.'' This can be easily achieved by introducing a fictitious spacetime dependent potential acting between the ''worm'' ends, so that their relative positions are now distributed according to the function

$$\mathcal{G}(i,\tau)\mathcal{Q}(i,\tau),$$

where $\mathcal{Q}(i,\tau)$ is arbitrary. In this way, one can change the typical size and shape of the loops generated by the ''worm'' algorithm.

The scope of the present paper is such that we are unable to discuss here many important details concerning the practical implementation of our algorithm (optimal triple-linked storage, particular forms of Eqs. (9), (10) for each subprocess, optimal management of subprocesses, etc.). Readers interested in these issues are encouraged to take advantage of our FORTRAN code with comments.

4. ILLUSTRATIVE RESULTS

To demonstrate the advantages of the CTWL algorithm, we have calculated properties of the 1-*D* boson Hubbard model (Eq. (1) with $U_{ij} = U_0\delta_{ij}$) for various coupling constants U_0 and particle densities ρ.

Comparison with the exact diagonalization results for small systems has demonstrated the lack of any detectable systematic error. In particular, for a system with eight lattice sites and six bosons, and on-site repulsion $U = 0.5$, the exact diagonalization result for the ground-state energy is $E_G = -10.49209$, while long-run Monte Carlo simulations yield $E_G = -10.4922(2)$, i.e., a result with relative accuracy better than 10^{-4}.

It is well known that a commensurate system with $\rho = 1$ undergoes a superfluid–Mott-insulator transition of Berezinskii–Kosterlitz–Thouless[23,24] type when the on-site interaction is strong enough (for the most accurate estimate of the transition point $U_0 = 1.645t$, see Ref. 25). In the superfluid phase, including the critical point, one can utilize knowledge of the long-wavelength behavior of the system. As explained by Haldane,[26] the energy associated with extra particles and nonzero winding numbers is quadratic in M and $N - \bar{N}$ (for simplicity, in what follows we count particle numbers from the commensurate value: $N \to N - L$ and $\bar{N} \to \bar{N} - L$). This means that the corresponding probability distribution in M and N is a Gaussian, i.e.,

$$W(N,M)\propto\exp\left[-\frac{L}{2\beta\Lambda_s(0)}M^2-\frac{\beta}{2L\kappa(0)}(N-\overline{N})^2\right]$$
$$\propto\exp\left[-\frac{\pi K(0)}{2}\left(\frac{L}{c\beta}M^2-\frac{c\beta}{L}(N-\overline{N})^2\right)\right]. \tag{21}$$

The zero argument of the superfluid stiffness Λ_s and compressibility κ denotes values at $T=0$. Here $K^{-1}=\pi\sqrt{\Lambda_s\kappa}$ is the index that controls the asymptotic behavior of the correlation functions, and c is the speed of sound.

At the critical point, $K(L)$, $\kappa(L)$ and $\Lambda_s(L)$ are system-size dependent quantities, with $K(L\to\infty)\to 1/2$. Since the speed of sound is unrenormalizable in a homogeneous system, it is sufficient to study scaling equations for the critical index only. In fact, the solution of the renormalization group (RG) equations for $K(L)$ can be ''visualized'' by considering the logarithmic derivative of the Green's function, since its index is just $K/2$:

$$K(l)=-2\frac{d\ \ln\mathcal{G}(r)}{d\ \ln r},\quad l=\ln r. \tag{22}$$

Here we have introduced the variable $r^2=x^2+(ct)^2$, which by conformal invariance describes asymptotic decay of $\mathcal{G}$ both in space and time.

Expressions (21) and (22) allow for a comprehensive test of the new algorithm. It is also tempting to consider a large system right at the quantum critical point and to evaluate its properties under the most unfavorable conditions for the standard worldline method. To calculate the critical index and the speed of sound, we considered a ring with 100 lattice sites and $\beta=100/t$. The critical parameters of the Hamiltonian are $U_0=1.645t$ and $\mu=1.94t$.[25] We had no problems in accumulating sufficient statistics of winding numbers and N for this system (the corresponding calculation is virtually impossible using the standard worldline algorithm). Simple manipulations with the exponents in (21) result in the following expressions:

$$\overline{N}=\frac{N^2-r_N}{2(N^2+r_N)},\quad r_N=\frac{\ln[W(0,N)/W(0,0)]}{\ln[W(0,-N)/W(0,0)]},$$
$$\kappa(0)=\frac{\beta}{L}\frac{N^2}{p_N},\quad p_N=-\ln\left[\frac{W(0,N)W(0,-N)}{W^2(0,0)}\right],$$
$$\Lambda_s(0)=\frac{L}{\beta}\frac{M^2}{g_M},\quad g_M=-\ln\left[\frac{W(0,M)W(0,-M)}{W^2(0,0)}\right]. \tag{23}$$

If one is interested in evaluating directly $K(0)$ then

$$K(0)=\frac{(p_Ng_M)^{1/2}}{\pi|NM|}. \tag{24}$$

The choice of N and M here is arbitrary, but for numeric reasons, the optimal N and M correspond to values where $(p_N,g_M)\sim 1$. The advantage of working with nonzero winding numbers in the grand canonical ensemble is obvious: in a single MC calculation, one collects all the necessary information about the parameters in the effective long-wavelength action, which is very convenient in determining quantum critical points from $K=K_c$. For the aforementioned system we found $c/t=2.4(1)$, and $K(l=\ln(100))=0.47(1)$.

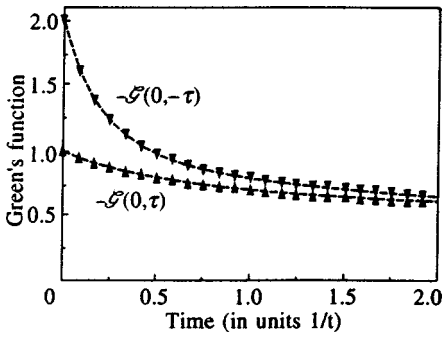


FIG. 4. Short-time behavior of the Green's function $\mathcal{G}(0,\tau)$ of the commensurate 1-D Hubbard model at the quantum critical point.

One note is in order here. The Gaussian distribution (21) implies that the system is in the superfluid phase. In the general case one has to define the compressibility as $\kappa=d\rho/d\mu$, where by definition $\rho=\overline{N}/L$. The superfluid stiffness Λ_s is defined as the coefficient relating persistent current and gauge phase when $\varphi\to 0$; this yields[27] $\Lambda_s=\overline{M}^2L/\beta$.

Finally, we used our method to evaluate the Green's function $\mathcal{G}(i,\tau)$ and to extract the critical index of the Berezinskii–Kosterlitz–Thouless transition from its asymptotic behavior; one can then check the consistency of all calculations. Since the CTWL simulation yields a two-dimensional histogram for $\mathcal{G}(i,\tau)$, much more accurate results for $K(l)$ are obtained by computing logarithmic derivatives along different directions in the (x,τ) plane with subsequent angular averaging. The speed of sound, which is necessary for such a calculation, is extracted from the asymmetry between x and τ in the asymptotic decay of $\mathcal{G}$. For the Green's function calculation, we considered a ring with 450 lattice sites and $\beta=200/t$. In Fig. 4, we show the short-range behavior of the Green's function in the τ-direction, with the characteristic jump at $\tau=0$. The dashed curve is a linear interpolation between the calculated points. In Fig. 5, we present the full-scale behavior of $\mathcal{G}(x,\tau)$ by plotting it as a function of $r=(x^2+(ct)^2)^{1/2}$ along the time and $x=c\tau$ directions. In accordance with conformal invariance, for large r the two curves are indistinguishable to within the statistical

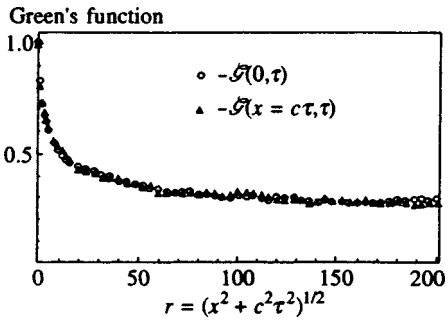


FIG. 5. Long-range behavior of $\mathcal{G}(i,\tau)$, demonstrating conformal invariance.

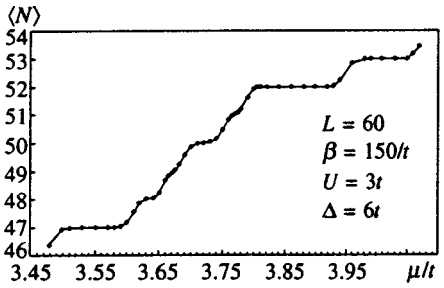


FIG. 6. Number of particles vs. chemical potential in a large Bose glass cluster at macroscopically low temperature.

errors. The speed of sound obtained from the Green's function is $c/t=2.4(1)$, and analysis of the logarithmic derivative (22) yields $K(l=\ln(100))=0.46(2)$.

It is worth noting that our calculations for $W(N,M)$ and $\mathscr{G}$ were performed on a Pentium-90 PC. None of these results (e.g., for $L>100$) can be obtained by other methods, even with the use of supercomputers.

The strong on-site disorder at low temperatures is a severe trial for most Monte Carlo schemes. Cluster methods suffer from inefficient global Metropolis updates here, while standard canonical-ensemble algorithms suffer from slowing down due to one-particle local minima in the effective action (the lowest single-particle states are well localized, and probing different configurations requires deep sub-barrier motion). The unique feature of our ''worm'' update method—the possibility of locally seeding an extra world line at any point in the spacetime continuum—obviates these problems.

To demonstrate the efficiency of our method, we present the results of just one nontrivial calculation—the dependence of the average particle number on the chemical potential in the Bose glass (BG) phase of the 1-*D* disordered Hubbard model, Fig. 6. We consider a system with $L=60$ sites at $\beta=150$, $U=3t$. Disorder is introduced by randomly distributing the on-site potential $\Sigma_i \epsilon_i n_i$ between $-\Delta$ and Δ, with $\Delta=6t$. The curve $\langle N\rangle(\mu)$ allows precise determination of the low-energy quasiparticle spectrum of the system to order $0.01t$, which is equivalent to the calculation of the total system energy to a relative accuracy of order 10^{-4}. The entire plot of $\langle N\rangle(\mu)$ was obtained in a few days of CPU time on a Pentium PRO-200 processor.

To obtain further evidence of the effectiveness of our method in more familiar problems, the reader is referred to a calculation of the superfluid—Bose glass–Mott insulator phase diagram in the 1-*D* disordered boson Hubbard model.[28]

5. CONCLUDING REMARKS

Although the CTWL algorithm developed here is quite general, some aspects deserve special discussion. What if the interaction radius r_0 is large? All of the procedures update configuration fragments with the typical duration $\tau_r-\tau_l \sim 1/t$. Since the method is exact, we trace all the kinks within the interaction radius, because they contribute to the function $R(\tau)$. This means that the interval (τ_l,τ_r) is further split into $N_{\tau_{l,r}}\sim r_0^d z \gg 1$ subintervals (z is the coordination number), and each subinterval requires special consideration. If r_0 is as large as the system size, then the whole scheme is in trouble, becoming a ''victim of exactness.''

The idea of solving such a problem is demonstrated by the stochastic series expansion method.[1,9,10] One might well wonder why continuous-time schemes, which contain as an essential ingredient an evaluation of time integrals, work as efficiently as SSE, which has all these integrals being evaluated exactly right at the start? Also, why is keeping the potential energy U in the τ-exponent not at all an advantage if $U\sim t$? The point is that the MC process is exact only for asymptotically long computation times, and there is no reason to calculate anything more precisely than the unavoidable statistical error, especially if the corresponding calculation becomes the bottleneck for the whole scheme. Evaluating time integrals in CTWL or reproducing exponents by expanding in power series in $U\sim t$ are just two cases that illustrate this point.

Suppose that all particles in the system interact with one another, so that formally $r_0=L$, but $\int d\mathbf{r}\rho(\mathbf{r})U(\mathbf{r})=F(L)\neq\infty$. We divide the interaction Hamiltonian into two parts, $H_{\rm int}^{(1)}(r<r_t)+H_{\rm int}^{(2)}(r>r_t)$, by introducing the truncation radius

$$\int_{|r|>r_t} d\mathbf{r}\,\bar{\rho}U(\mathbf{r})=t. \tag{25}$$

We then write $H_0=H_{\rm int}^{(1)}$ and combine $H_{\rm int}^{(2)}$ with V (see Sec. 2), i.e., the long-range part of the interaction Hamiltonian is now considered to consist of diagonal kinks. Because of the definition (25), the total number of kinks within the time interval $\sim 1/t$ remains finite and independent of system size.

The case of a divergent integral $\int d\mathbf{r}\,\rho(\mathbf{r})U(\mathbf{r})=F(L)\to\infty$ is more subtle, since the number of diagonal kinks within the time interval $\sim 1/t$, given by $F(L)$, is now large (if $F(L)$ is a logarithmic function of L, we do not regard this problem as serious). On the other hand, for long-range interactions the so called ''mean-field approximation'' becomes more accurate. Since the mean-field potential is easy to account for analytically (and numerically), one now has to deal with fluctuations, and these quite often satisfy the condition

$$\left|\int d\mathbf{r}\,(\rho(\mathbf{r})-\bar{\rho})U(\mathbf{r})\right|=\delta F(L)\neq\infty. \tag{26}$$

In Appendix A we explain how to organize the Monte Carlo process using the mean-field approximation for the configuration weight. The net result is that even for long-range potentials, the calculation time can remain independent of system size.

In this paper we have concentrated on the Green's function calculation by restricting the number of worldline discontinuities to 0 or 2. Of course the scheme can be trivially extended to include the case with a larger number of discontinuities, if one is interested in the two- or *n*-particle Green's function or *n*-point vertex. More generally, our scheme makes it possible to work with Hamiltonians that do not conserve the number of particles, i.e., when there are sources with finite strength in the bare Hamiltonian.

Although in this paper we consider a system with discrete Hilbert space in detail, the principles of update in continuous time developed here are much more general. Mathematically, we construct an exact method (in the statistical limit) of averaging over a distribution represented as a series of integrals with an ever-increasing number of variables, but with essential similarity among the terms of the series, allowing their local comparison (weighting). We may call such structures integrals with a variable number of variables—VNV integrals. Physically, we sum a perturbative expansion in the interaction picture for some observable of a large but essentially finite-size system. (For a system with discrete Hilbert space, the only continuous variables in this expansion are the times of virtual transitions.) But perturbative expansions for continuous systems also have the structure of VNV integrals, with additional integrations over some continuous variables. Thus (apart from the fact that for spatially continuous systems one cannot expand the kinetic part of the Hamiltonian and must use the potential energy as a perturbation), there is no qualitative difference between perturbative expansions for continuous and discrete systems. The general method of evaluating VNV integrals is given by Eqs. (6) and (8)–(10), where the vector $\boldsymbol{\tau}$ now stands for any set of continuous variables, and the function $R(\boldsymbol{\tau})$ is defined straightforwardly, given the particular form of the series.

We would like to thank V. Kashurnikov, A. Sandvik, M. Troyer, H. Evertz, B. Beard, and N. Kawashima for inspiring discussions of existing Monte Carlo schemes and valuable comments on the final version of the paper. This work was supported by Grant No. INTAS-93-2834-ext (of the European Community) and partially by the Russian Fund for Fundamental Research (Grant No. 95-02-06191a).

APPENDIX

Long-range potentials

Suppose that we are dealing with the case $F(L)\to\infty$, but finite $\delta F(L)$. The idea is to organize the Monte Carlo process in such a way, that in most updates we simply ignore fluctuations, and account for distant particles by replacing them with a homogeneous density distribution. Obviously, for the scheme to remain accurate, in some updates we have to consider deviations from the mean-field distribution. The goal is to address the procedure dealing with distant fluctuations with the small probability which is at least inversely proportional to the number of operations in this procedure.

Consider again the balance equation for the given pair of subprocesses, but now including the possibility of completing the same update procedure in a number of ways:

$$A_0\, p_c W(\boldsymbol{\tau}) \sum_{j=0}^{j_*} \gamma^{(j)} P^{(j)}_{acc}(\boldsymbol{\tau})\, d\boldsymbol{\tau} - dA_n(\boldsymbol{\tau}) p_a \times \sum_{j=0}^{j_*} \gamma^{(j)} P^{(j)}_{rem}(\boldsymbol{\tau}) = 0. \tag{A1}$$

Here $\gamma^{(j)}$ is the probability of using the jth version of the update procedure. We require

$$\sum_{j=0}^{j_*} \gamma^{(j)} = 1, \quad \text{and} \quad \gamma_0 \gg \gamma_1 \ldots \gg \gamma_{j_*}.$$

We also assume that the procedure j_* corresponds to the exact treatment of all fluctuations. Other quantities have exactly the same meaning as in (6). The self-balance condition now reads (compare Eq. (8))

$$W(\boldsymbol{\tau}) \sum_{j=0}^{j_*} \gamma^{(j)} P^{(j)}_{acc}(\boldsymbol{\tau}) = R(\boldsymbol{\tau}) \sum_{j=0}^{j_*} \gamma^{(j)} P^{(j)}_{rem}(\boldsymbol{\tau}). \tag{A2}$$

To satisfy (A2) we suggest the following scheme. Let $R^{(j)}(\boldsymbol{\tau})$ be the distribution corresponding to the exact treatment of fluctuations up to the distance $r^{(j)}$ with $r^{(0)} \ll r^{(1)} \ll \ldots \ll r^{(j_*)} = L$, and the mean-field treatment of more distant ($r > r^{(j)}$) particles. We can write then

$$R^{(j)} = R^{(0)} + \delta R^{(1)} + \ldots \delta R^{(j)}, \quad R^{(j_*)} \equiv R. \tag{A3}$$

If δF is finite and $r^{(0)}$ is sufficiently large, then all $\delta R^{(j)}$ are small. We then choose $\gamma^{(0)} \approx 1$ and

$$P^{(0)}_{acc}(\boldsymbol{\tau}) = \begin{cases} R^{(0)}(\boldsymbol{\tau})/W(\boldsymbol{\tau}), & \text{if } R^{(0)}(\boldsymbol{\tau}) < W(\boldsymbol{\tau}) \\ 1, & \text{otherwise} \end{cases}, \tag{A4}$$

$$P^{(0)}_{rem}(\boldsymbol{\tau}) = \begin{cases} W(\boldsymbol{\tau})/R^{(0)}(\boldsymbol{\tau}), & \text{if } R^{(0)}(\boldsymbol{\tau}) < W(\boldsymbol{\tau}) \\ 1, & \text{otherwise} \end{cases}, \tag{A5}$$

and solve the self-balance condition deductively by requiring

$$W(\boldsymbol{\tau}) \sum_{j=0}^{k} \gamma^{(j)} P^{(j)}_{acc}(\boldsymbol{\tau}) = R^{(k)}(\boldsymbol{\tau}) \sum_{j=0}^{k} \gamma^{(j)} P^{(j)}_{rem}(\boldsymbol{\tau}), \tag{A6}$$

or equivalently

$$\gamma^{(k)} [W P^{(k)}_{acc} - R^{(k)} P^{(k)}_{rem}] = \delta R^{(k)} \sum_{j=0}^{k-1} \gamma^{(j)} P^{(j)}_{rem}. \tag{A7}$$

The final answer can be written

$$P^{(k)}_{acc} = \begin{cases} \left[\delta R^{(k)} \sum_{j=0}^{k-1} \gamma^{(j)} P^{(j)}_{rem} \right] \Big/ [\gamma^{(k)} W(\boldsymbol{\tau})], & \text{if } \delta R^{(k)} > 0 \\ 0, & \text{otherwise} \end{cases}, \tag{A8}$$

$$P^{(k)}_{rem} = \begin{cases} -\left[\delta R^{(k)} \sum_{j=0}^{k-1} \gamma^{(j)} P^{(j)}_{rem} \right] \Big/ [\gamma^{(k)} R^{(k)}(\boldsymbol{\tau})], & \text{if } \delta R^{(k)} < 0 \\ 0, & \text{otherwise} \end{cases}. \tag{A9}$$

Since all $\delta R^{(k)}$ are assumed to be small, it is possible to keep $\gamma^{(k)} \ll 1$ (for $k = 1,2,\ldots,j_*$), but large enough to avoid situations with $P^{(k)}_{acc} > 1$ or $P^{(k)}_{rem} > 1$.

*)E-mail: prokofev@kurm.polyn.kiae.su

1)To stress this important point, we find it reasonable to distinguish between ''efficient'' LCU algorithms and others. By ''efficient'' LCU we mean

algorithms in which detailed balance is taken care of in the cluster-building rules, not in having a global Metropolis step with small acceptance rates for large clusters. Nevertheless, in certain cases ''inefficient'' LCU code works reasonably well.[15]

[1] A. W. Sandvik, To appear in Proc. of the 1996 El Escorial Summer School on Strongly Correlated Magnetic and Superconducting Systems (Springer Verlag).
[2] M. Suzuki, Prog. Theor. Phys. **56**, 1454 (1976); M. Suzuki, S. Miyashita, and A. Kuroda, Prog. Theor. Phys. **58**, 1377 (1977).
[3] J. E. Hirsch, D. J. Scalapino, R. L. Sugar, and R. Blankenbecler, Phys. Rev. Lett. **47**, 1628 (1981); J. E. Hirsch, R. L. Sugar, D. J. Scalapino, and R. Blankenbecler, Phys. Rev. B **26**, 5033 (1982).
[4] V. A. Kashurnikov and Yu. G. Kharchenko, JETP Lett. **63**, 900 (1996).
[5] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Phys. Rev. B **24**, 2278 (1981); D. J. Scalapino and R. L. Sugar, Phys. Rev. B **24**, 4295 (1981).
[6] J. E. Hirsch, Phys. Rev. B **31**, 4403 (1985).
[7] S. R. White, D. J. Scalapino, and R. L. Sugar, Phys. Rev. Lett. **63**, 1523 (1989); S. R. White, D. J. Scalapino, R. L. Sugar, E. Y. Loh, J. E. Gubernatis, and R. T. Scalettear, Phys. Rev. B **40**, 506 (1989); R. Preuss, W. Hanke, W. von der Linden, Phys. Rev. Lett. **75**, 1344 (1995).
[8] N. Travedi and D. M. Ceperley, Phys. Rev. B **41**, 4552 (1990); S. Sorella, A. Parora, M. Parrinello, and E. Tosatti, Europhys. Lett. **12**, 721 (1990).
[9] A. W. Sandvik and J. Kurkijärvi, Phys. Rev. B **43**, 5950 (1991).
[10] A. W. Sandvik, J. Phys. A **25**, 3667 (1992).
[11] D. C. Handscomb, Proc. Cambridge Philos. Soc. **58**, 594 (1962); Proc. Cambridge Philos. Soc. **60**, 115 (1964).
[12] R.-H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
[13] H. G. Evertz, G. Lana, and M. Marcu, Phys. Rev. Lett. **70**, 875 (1993).
[14] U.-J. Wiese and H.-P. Ying, Z. Phys. B **93**, 147 (1994).
[15] N. Kawashima, J. E. Gubernatis, and H. G. Evertz, Phys. Rev. B **50**, 136 (1994).
[16] N. Kawashima and J. E. Gubernatis, Phys. Rev. Lett. **73**, 1295 (1994); J. Stat. Phys. **80**, 169 (1995).
[17] N. Kawashima, J. Stat. Phys. **82**, 131 (1996).
[18] B. B. Beard and U.-J. Wiese, Phys. Rev. Lett. **77**, 5130 (1996).
[19] E. Farhi and S. Gutmann, Ann. Phys. (N.Y.) **213**, 182 (1992).
[20] N. V. Prokof'ev, B. V. Svistunov, and I. S. Tupitsyn, JETP Lett. **64**, 911 (1996).
[21] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth *et al.*, J. Chem. Phys. **21**, 1087 (1953).
[22] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Quantum Field Theory Methods in Statistical Physics*, Nauka, Moscow (1962).
[23] V. L. Berezinskiĭ, Zh. Éksp. Teor. Fiz. 61, 1144 (1971) [Sov. Phys. JETP **64**, 610 (1972)].
[24] J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973); J. M. Kosterlitz, J. Phys. C **7**, 1046 (1974).
[25] V. A. Kashurnikov and B. V. Svistunov, Phys. Rev. B **53**, 11776 (1996); V. A. Kashurnikov, A. V. Krasavin, and B. V. Svistunov, JETP Lett. **64**, 104 (1996).
[26] F. D. M. Haldane, Phys. Rev. Lett. **47**, 1840 (1981); J. Phys. C **14**, 2585 (1981).
[27] E. L. Pollock and D. M. Ceperley, Phys. Rev. B **36**, 8343 (1987).
[28] N. V. Prokof'ev and B. V. Svistunov, submitted to Phys. Rev. Lett., preprint cond-mat/9706169.

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

Josephson array of mesoscopic objects. Modulation of system properties through the chemical potential

A. I. Belousov, S. A. Berzakov, and Yu. E. Lozovik*)

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Moscow
(Submitted 8 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 591–604 (August 1998)

The phase diagram of a two-dimensional Josephson array of mesoscopic objects (superconducting granules, superfluid helium in a porous medium, traps with Bose-condensed atoms, etc.) is examined. Quantum fluctuations in both the modulus and phase of the superconducting order parameter are taken into account within a lattice boson Hubbard model. Modulating the average occupation number n_0 of the sites in the system (the ''number of Cooper pairs'' per granule, the number of atoms in a trap, etc.) leads to changes in the state of the array, and the character of these changes depends significantly on the region of the phase diagram being examined. In the region where there are large quantum fluctuations in the phase of the superconducting order parameter, variation of the chemical potential causes oscillations with alternating superconducting (superfluid) and normal states of the array. On the other hand, in the region where the bosons interact weakly, the properties of the system depend monotonically on n_0. Lowering the temperature and increasing the particle interaction force lead to a reduction in the width of the region of variation in n_0 within which the system properties depend weakly on the average occupation number. The phase diagram of the array is obtained by mapping this quantum system onto a classical two-dimensional XY model with a renormalized Josephson coupling constant and is consistent with our quantum path-integral Monte Carlo calculations. © *1998 American Institute of Physics.*
[S1063-7761(98)01608-4]

1. INTRODUCTION

Advances in microlithographic techniques have made it possible to create regular arrays of mesoscopic Josephson objects. These systems, which are under active experimental and theoretical study, include superfluid helium in porous media[1)] (see Ref. 2 and the literature cited there), lattices of mesoscopic Josephson contacts,[3,4] and ultrasmall superconducting granules.[5,6] One interesting physical realization of a Josephson array consists of Josephson junctions in a structure with superfluid ^{3}He created by means of lithography.[7] Major advances in experiments with Bose condensates of atoms cooled by laser irradiation followed by evaporation[8–10] suggest that it may be possible to fabricate a Josephson array of close magneto-optical traps with Bose-condensed atoms[2)] or with clusters of Bose-condensed atoms cooled and localized at the nodes of a system of standing electromagnetic waves. Finally, another remarkable realization of a Josephson array might be a system of Josephson coupled ''lakes'' of Bose-condensed excitons in single or double quantum wells located at the minima of the random field created by the roughness of the well surfaces, i.e., in ''natural'' quantum dots,[12] or in an array of artificial quantum dots.

For concreteness, the following discussion considers the example of a system of superconducting mesoscopic granules or Josephson junctions, but the results apply to all the systems mentioned above. We shall consider a regular array of mesoscopic granules situated on a conducting substrate and separated from it by a thin dielectric layer. A voltage applied to the conducting substrate serves as the chemical potential of the Cooper pairs, which determines the average occupation number n_0 of the granules in the system.[13,14] For example, in the case of superfluid helium in a porous medium, the chemical potential of the atoms can be varied (because of the contribution of the van der Waals interaction) by changing the thickness of the layer of adsorbed helium.[15,16] The state of an array of mesoscopic traps with cooled atoms can be controlled by changing the average number of atoms in the system (by, for example, capturing them from an external flow).

The systems being examined here, in which the character of the processes taking place is determined by the boson degrees of freedom, are conveniently described by a lattice boson Hubbard model[17,18] with the Hamiltonian

$$\hat{H}=\frac{t}{2}\sum_{\langle i,j\rangle}(2a_i^\dagger a_i-a_i^\dagger a_j-a_j^\dagger a_i)+\frac{U}{2}\sum_i(a_i^\dagger a_i)^2-\mu\sum_i a_i^\dagger a_i. \quad (1)$$

In this model site i corresponds to one superconducting granule or pore with helium, to a single trap with a Bose condensate, etc. The operators $a_i^\dagger$ (a_i) are the Bose creation (annihilation) operators of an ''effective'' boson at site $i=\overline{1,N^2}$ of an $N\times N$ lattice. The first term in the Hamiltonian takes into account the kinetic energy of the particles, which corre-

sponds to the Josephson tunneling energy and is described by the parameter t. The second term in Eq. (1) describes the interaction of effective bosons at a granule with a characteristic energy $U_i>0$.

The model (1) is interesting in that it can be used to study the properties of arrays of mesoscopic structures in which the relative fluctuations in the modulus of the superconducting order parameter are large. In this regard, we note that the quantum XY model is justified only when these fluctuations are small,[14] i.e., in the case of arrays of macroscopic granules.

A lattice system (1) with Mott-insulator and superconducting phases at $T=0$ has been studied both analytically[17–19] and by computer simulation.[20,21] In this paper we shall be interested in the properties of the system (1) at finite temperatures, digressing from the interesting phase transitions at $T=0$.[22] In some earlier papers,[23] the investigation was limited to the case of integer (commensurable) populating, in which the average number of bosons per site (granule), $n_0=\langle a_i^\dagger a_i\rangle$, is a whole number. At $T=0$, adding even a single particle to an arbitrarily large system changes its properties in a fundamental way. Specifically, a system with a noninteger average number of bosons per granule remains superconducting for arbitrary values of U/t for the interaction between the particles.[17–22] It is evident that this surprising behavior should be present even in the limit of $T=0$. In fact, as will be shown below, at finite temperatures the properties of the system vary little within an interval $n_0=k\pm\delta n_0$ about an integer population, whose width $2\delta n_0$ decreases as the temperature is lowered.

The purpose of this paper is to study the changes in the character of the ordering in an array of granules as the substrate voltage (chemical potential of the effective bosons) is varied. Here we shall not use the simplifying assumption of small relative fluctuations in the modulus of the superconducting order parameter. The results given below correspond to an array of mesoscopic objects for which the root-mean-square fluctuations in the number of particles are comparable to their average number. In Sec. 2 we present results from mean-field calculations. The method used there corresponds to mapping the initial boson model (1) onto an effective classical XY model with a renormalized Josephson coupling constant. To refine the results of the analytical calculations, we use the quantum path-integral Monte Carlo method (see Sec. 3). A discussion and comparison of the results in Sec. 4 completes the presentation.

2. BOSON HUBBARD MODEL IN THE MEAN-FIELD APPROXIMATION

A qualitative approximation for the phase diagram of the model (1) can be obtained using an approach described previously.[19,23,24] In terms of this model, the boundary of the ordered state is located where the local density of the superconducting component vanishes in the effective functional describing long-wavelength excitations of the system. The latter can be obtained in the usual way using the Hubbard–Stratonovich transform[25] followed by an expansion of the effective functional in components of the fluctuating field.[19,24] The condition that the local density of the superconducting component vanish for the system described by the Hamiltonian (1) yields an equation for determining the boundary of the ordered state:

$$\tilde{q}^2=4\sum_{n=0}^{\infty}\left[\exp\left(-\frac{0.5\tilde{q}^2(n-\tilde{\eta})^2}{\tilde{T}}\right)-\exp\left(-\frac{0.5\tilde{q}^2(n+1-\tilde{\eta})^2}{\tilde{T}}\right)\right]\frac{n+1}{2(n-\tilde{\eta})+1}\times\left[\sum_{n=0}^{\infty}n\,\exp\left(-\frac{0.5\tilde{q}^2(n-\tilde{\eta})^2}{\tilde{T}}\right)\right]^{-1}. \tag{2}$$

Here we have used the following independent dimensionless parameters, which specify the state of the system:

$$\tilde{q}=\sqrt{\frac{U}{t}},\quad \tilde{T}=\frac{k_BT}{t},\quad \tilde{\eta}=\frac{\mu}{U}-\frac{2t}{U}.$$

The average number of particles in the granules of the system in the disordered state is given by the equation

$$n_0=\frac{\Sigma_{n=0}^{\infty}n\,\exp\{-0.5\tilde{q}^2(n-\tilde{\eta})^2/\tilde{T}\}}{\Sigma_{n=0}^{\infty}\exp\{-0.5\tilde{q}^2(n-\tilde{\eta})^2/\tilde{T}\}}. \tag{3}$$

The solid curves in Fig. 1a represent the phase diagram of the system (1) in the variables $\{\sqrt{U/t},\mu/U\}$ obtained by solving Eq. (2) for different temperatures k_BT/t. The region of low values of U/t (low particle interaction energies) corresponds to the superconducting (S) state of the system.

As $k_BT/t\to0$, the disordered state of the system corresponds to an integer average number of particles per granule, $n_0=k$, and determines the domain of existence of the Mott insulator (I).[17] In this limit, the superfluid state of the system corresponds to the case of incommensurate populations, i.e., to a noninteger average number of particles per granule. Figure 1a shows that for $T=0$ and half-integer values of the chemical potential $\mu=0.5+k$, a superconducting state exists for arbitrarily strong interparticle interactions, in accord with earlier work.[17,18]

At finite temperatures an increase in the boson interaction force leads the system into a disordered state for arbitrary values of the chemical potential (see Fig. 1a). It is also clear from the figure that as the temperature is increased, the domain of existence of the ordered state is shifted toward higher values of the chemical potential.

The transition to the case of a system of macroscopic granules corresponds to increasing the particle density n_0 and reducing the role of the fluctuations in the modulus of the order parameter. It is most convenient to follow the changes in system (1) with increasing n_0 in the $\{q,\Theta\}$ plane, where we use the dimensionless temperature $\Theta=k_BT/tn_0$ and the quantum parameter $q=\sqrt{U/tn_0}$, which are the control parameters that also determine the state of the quantum XY model. The corresponding phase diagram is shown in Fig. 2a. As can be seen from this figure, for any values of U the estimate of the boundary of the ordered state in the Hubbard model according to mean-field theory lies above the corresponding limit in the XY model and approaches it as the

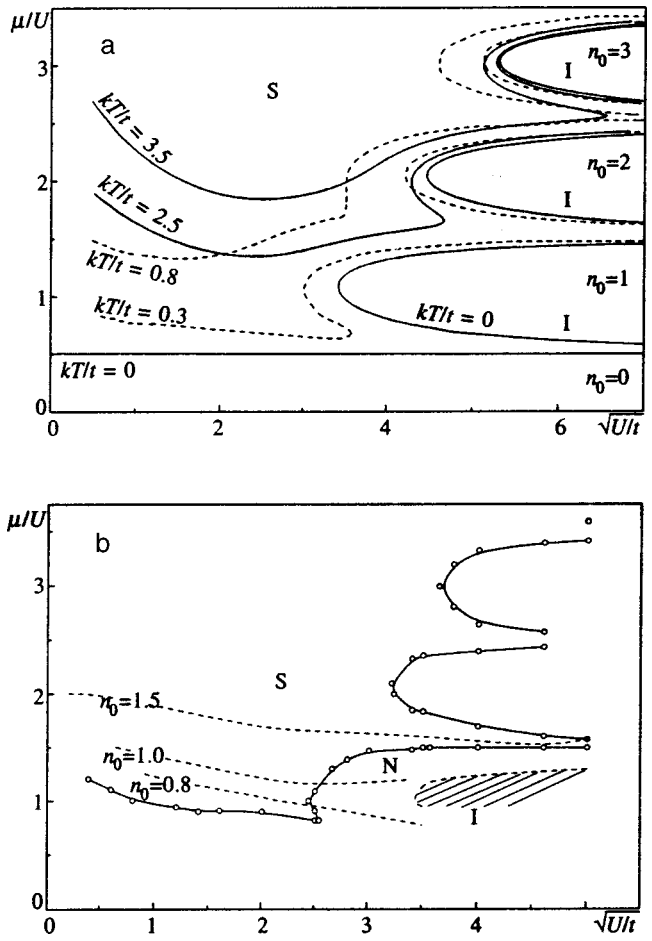


FIG. 1. Phase diagram for the Hubbard model (1) in the coordinates $\{\sqrt{U/t},\mu/U\}$. S—the superconducting state; N—the normal (metallic) state; I—Mott insulator (hatched region in Fig. b). (a) Mean-field calculation. The solid curves were obtained by solving Eq. (2) and correspond to vanishing of the local density of the superfluid component. The Kosterlitz–Thouless topological transition (6) takes place on the dashed curves. (b) Monte Carlo calculations for $k_BT/t=0.8$.

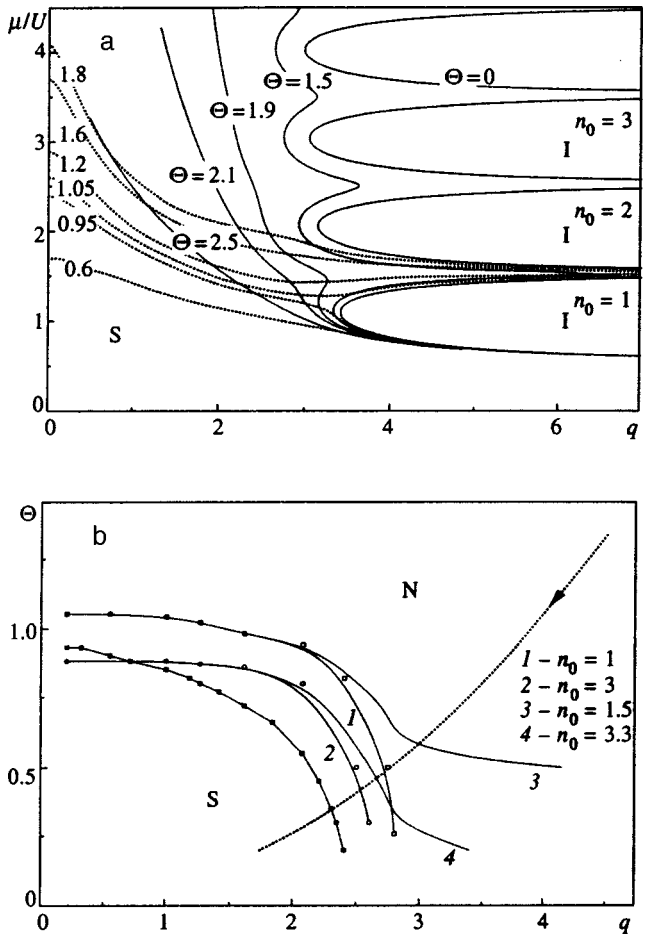


FIG. 2. Phase diagram for the Hubbard model (1) in the coordinates $\{q,\mu/U\}$ and $\{q,\Theta\}$. (a) Mean-field theory calculation. The solid curves were obtained by solving Eq. (2). The dotted curves correspond to the joint solution of the system of Eqs. (2) and (3). (b) Monte Carlo calculations. The filled squares are the phase diagram for the $2+1$-dimensional XY model (for integer values of $n=k\gg 1$).[34] The system with $\{U/t,k_BT/t\}=\{3.5,0.8\}$ and a variable chemical potential μ/U moves along the dotted curve (see Figs. 4 and 5 below).

average population n_0 of the lattice sites increases. Our calculations confirm that the phase diagram is periodic in the parameter μ/U when $n_0\gg 1$.[14]

The dotted lines in Fig. 2a comprise a family of curves, at whose points a system with density n_0 becomes disordered. The points where they intersect the $\Theta=\text{const}$ lines determine the phase diagram of a system in the $\{q,\Theta\}$ plane corresponding to n_0 particles per granule. A similar analysis shows that for an incommensurate boson density at low temperatures (see below), an ordered (superconducting) state of an array exists for arbitrarily large values of q, i.e., for arbitrarily large quantum fluctuations in the phase of the superconducting order parameter in terms of the quantum XY model.

This approach only yields a qualitative estimate of the characteristic features of the phase diagram of this system. A comparison with the results of a numerical simulation (see below) shows that Eqs. (2) and (3) give a greatly overestimated value for the disorder temperature $\Theta_c(q;\mu/U)$. In order to obtain more accurate quantitative estimates, it is necessary to determine the temperature at which the global (rather than local, as in the method described above) superfluid density of the array vanishes. This temperature can be estimated as the Kosterlitz–Thouless temperature for the topological phase transition according to the classical XY model, onto which the initial system is mapped by expanding the effective Ginzburg–Landau functional for weak fluctuations in the phase of the order parameter.

Using the approach in Refs. 19, 24, and 26 it is easy to show that the effective action of the classical two-dimensional XY model sought is given by

$$S(\{\varphi_{\mathbf{k}}\})=\frac{J_{XY}}{2}\sum_{\mathbf{k}}|\mathbf{k}|^2\varphi_{\mathbf{k}}\varphi_{-\mathbf{k}}\approx J_{XY}\times\sum_{\langle i,j\rangle}[1-\cos(\varphi_i-\varphi_j)],$$
$$J_{XY}\left(\frac{\mu}{U};\frac{t}{U};\frac{k_BT}{t}\right)=\frac{t\Delta^2}{4}, \tag{4}$$

where J_{XY} is the coupling constant for the effective XY model, which depends on the dimensionless parameters μ/U, t/U, and k_BT/t. The local superfluid density $\Delta^2/4$ of the system is described by the relation

$$\Delta=\frac{\mathrm{Tr}\{(\hat{a}^\dagger+\hat{a})\exp(-\beta\hat{H}_{mf})\}}{\mathrm{Tr}\{\exp(-\beta\hat{H}_{mf})\}},$$
$$\hat{H}_{mf}=\frac{U}{2}\hat{n}^2+(2t-\mu)\hat{n}-t\Delta(\hat{a}^\dagger+\hat{a}). \tag{5}$$

A plot of the Kosterlitz–Thouless phase-transition temperatures in the effective XY model[27] specifies the boundary sought of the superfluid state for this array of granules:

$$k_BT_c=0.98J_{XY}(\mu/U;t/U;k_BT_c/t). \tag{6}$$

Estimates given by Eqs. (4) and (5) are shown as dashed curves in Fig. 1a. Note that, although the two approaches discussed above yield a similar qualitative behavior of the boundary of the ordered phase, the temperature of the topological phase transition in the effective XY model (6) is considerably lower than the temperature at which the local superfluid density vanishes according to Eq. (2). A comparison of the phase diagram obtained in this way with the results of Monte Carlo calculations (see below and Fig. 1) shows that they are in fair quantitative agreement.

3. QUANTUM MONTE CARLO METHOD. MEASURABLE QUANTITIES

The Trotter discretization procedure makes it possible to estimate all the thermodynamic averages of the operators of a D-dimensional quantum system in terms of a classical $D+1$-dimensional system, where the product of the matrix elements (calculated approximately) of the high temperature density matrix serves as the Boltzmann weight of the configurations of the corresponding classical system. For studying the properties of the model (1), we shall use the ''checkerboard version'' of the quantum Monte Carlo method. (A detailed discussion of the discretization procedure and the organization of the Monte Carlo step during the simulation of systems of lattice bosons in a large canonical ensemble is given elsewhere.[28]) In this method, the degrees of freedom of the discretized system are the occupation numbers $\{n_i^p\}$ of the sites of the $N\times N\times 4P$ three-dimensional lattice formed by $4P$-fold multiplication of the initial $N\times N$ lattice along the imaginary time axis. The number of subdivisions P was chosen so that the parameter $\epsilon=q^2/P^2\Theta^2$, which characterizes the discretization error, would be less than 0.06.

The density ν_s of the superfluid component was calculated at each computational point of the phase diagram, whose position is specified by the parameters $\{\sqrt{U/t},k_BT/t\}$ and the chemical potential μ/U. To find this quantity, we used both the fluctuations in the topological winding number[20,28] and the correlation function of the paramagnetic current.[29] We found that when the average particle density at the boundary $n_0<2$ and $q>2$, the statistical errors in the second method were considerably higher than the errors in determining the superfluid density from the fluctuations in the winding number, rendering it unsuitable.

We also measured n_0, which is controlled by the chemical potential of the system, and the compressibility modulus κ, which is defined as

$$\kappa=k_BT\frac{\partial n_0}{\partial\mu}=\frac{1}{4PN^2}\left\langle\sum_{p=0}^{4P-1}\sum_i(n_i^p)^2\right\rangle-(n_0)^2. \tag{7}$$

It turned out to be convenient to make the measurements with $U/t=\mathrm{const}$ and $k_BT/t=\mathrm{const}$ and variation of the chemical potential μ/U. As an example, in Fig. 2b the line along which the system moves for $\{U/t,k_BT/t\}=\{3.5,0.8\}$ is plotted in the coordinates $\{q,\Theta\}$. The location of the system on this line for a given value of the chemical potential can be determined by measuring the average number n_0 of particles per granule. In addition, we did a number of calculations for fixed n_0 (in a canonical ensemble). It might be expected that for the same particle density n_0 the results of the simulation would be independent of the choice of the ensemble for sufficiently large systems. We tested this assumption and found for a system of dimension $N\times N=6\times 6$ that the difference in the measured quantities is less than 10% in the region of variation of the control parameters of interest to us. Therefore, in analyzing the results, data obtained by assigning the density n_0 (within a canonical distribution) and by assigning the chemical potential (which corresponds to using a grand canonical ensemble) can be used simultaneously.

4. DESCRIPTION AND DISCUSSION OF RESULTS

We first consider the computational results for $q<1.5$, where the interparticle interaction is weak and mean-field theory (see Figs. 1a and 2a) predicts a monotonic dependence of the phase-transition temperature on the particle density n_0. Figure 3 shows the calculated density of the superfluid component as a function of average occupation number, $\nu_s(n_0)$, for $q=1$ and $\Theta=1$ (unfilled symbols). As the occupation number increases, ν_s approaches a constant equal to the helicity modulus $\gamma(q,\Theta)$ in the $2+1$-dimensional (quantum) XY model.[30,31] We found previously[23] that in this region of the phase diagram, this limit, which corresponds to a small contribution of the quantum fluctuations to the modulus of the order parameter, is approached already when $n_0=4-5$. The monotonicity of $\nu_s(n_0)$ suggests that the results of experiments on a system of isolated granules will not differ greatly from those on a system of granules on a substrate with an applied potential. It turned out that the system behaves similarly up to $q\simeq 2.3$. As the quantum parameter q is raised further and the temperature Θ is lowered, $\nu_s(n_0)$ ceases to be a monotonic function of the average occupation number n_0. Characteristic oscillations in $\nu_s(n_0)$ with minima at integer values of $n_0=k$ are noticeable in Fig. 3 ($\{q,\Theta\}=\{2.75,0.5\}$, filled symbols). For sufficiently high boson densities ($n_0>7$ in the region $\{q,\Theta\}\approx\{2.5,0.5\}$; see Ref. 23) there is a transition to the quasiclassical limit and the density of the superconducting component $\nu_s(n_0)$ becomes a periodic function of the average occupation number with a period of unity.[14]

Figure 4a shows plots of the superfluid density as a function of the chemical potential of the system, which is proportional to the voltage applied to the substrate. The squares

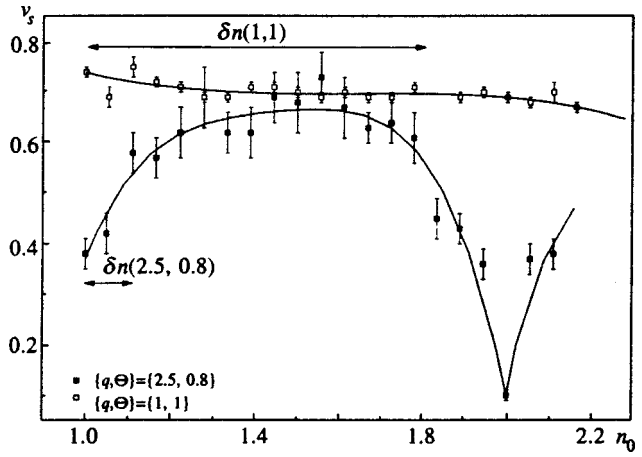


FIG. 3. Plots of the fraction ν_s of the superfluid component as a function of the average occupation number n_0. The solid curves were obtained by interpolation of the data with a fourth-order polynomial. The lengths of the horizontal arrows correspond to estimates of $\delta n(q,\Theta)=\kappa(q,\Theta)|_{n_0=1}$. See Eq. (8).

refer to a system with $\sqrt{U/t}=2$, $k_BT/t=0.8$ and the triangles to $\sqrt{U/t}=2.5$, $k_BT/t=0.8$. The feature at $n_0\simeq 1$ [a dip on the plots of $\nu_s(\mu/U)$ and $\kappa(\mu/U)$ for at $\mu/U\simeq 1.2$] is poorly seen for $q\simeq 2$ (squares in Fig. 4) but becomes clearly evident for $q\simeq 2.5$ (triangles). The figure confirms that the case of commensurate populations corresponds to lower densities of the superfluid component, i.e., increasing the deviation of the average boson density from integer values of $n_0=k$ leads to spreading of the phase diagrams for the model (1) in the $\{q,\Theta\}$ plane.

As the quantum parameter q is increased further, the differences in the properties of the system for integer and noninteger boson densities become increasingly more significant. In fact, based on the results from mean-field calculations (see Fig. 1a), we can assume that there is a value of U/t for which the $\sqrt{U/t}=$const line intersects the region of the disordered state with $n_0\simeq 1$. Further increases in the interaction constant U should lead to the possibility of an intersection with the disordered region having $n_0\simeq 2$, etc. The computational results shown in Fig. 5 confirm this assumption. Note that for $\sqrt{U/t}=5.0$ (triangles in Fig. 5) and integer values of μ/U, changes in the chemical potential lead to essentially no change in the average number of particles at the sites in the system. This feature, which is characteristic of an insulator, can also be seen in Fig. 5b, which shows a plot of the compressibility modulus κ as a function of the chemical potential μ/U.

At $T=0$ a lattice boson system without disorder undergoes a superconductor-insulator phase transition.[17] As for the system being studied here, we may assume that as the temperature is raised, the domain of existence of the Mott insulator decreases and shifts toward larger U/t, so that, for example, on moving along the $\mu/U=k$ line the superfluid phase is replaced by the normal (metallic) phase. Further increases in the interaction constant, under which the magnitude of the Mott dielectric gap increases so much that thermal excitations become unimportant, lead to crossover formation of the insulator state.[32] An estimate of the location of the boundary of the insulator state for $n_0=1$, based on our calculations, is shown in Fig. 1b. Several constant-density ($n_0=$const) contours are also shown for comparison. A more detailed study of crossover-type metal-insulator transitions requires that systems of substantially larger size be examined.[24,33,34]

An analysis of Figs. 1–5 shows that interesting effects caused by incommensurate populating of the sites in the system occur only in quite strongly interacting systems and at sufficiently low temperatures. For the boson Hubbard model (1), the domain in which they exist can be estimated by the inequalities $q>2.3$ and $\Theta<0.7$. Quantitatively, the magnitude of the deviation δn of the average particle density from an integer value of $n_0=k$ at which a significant change in the system properties will be observed can be expressed in terms of the compressibility modulus:

$$\delta n\approx\left.\frac{\partial n_0}{\partial\mu}\right|_{n_0=k} k_BT=\kappa|_{n_0=k}. \tag{8}$$

It is known that the compressibility modulus [which is inversely proportional to the phase fluctuations in the $2+1$-dimensional XY model; see Eq. (7)] falls off substantially as the quantum parameter q becomes greater and the temperature Θ is reduced. We found that the following estimates hold for the Hubbard model (1) when $n_0=1$: $\delta n\approx 1.2$ for $\{q,\Theta\}\approx\{0.5,1\}$, $\delta n\approx 0.3$ for $\{q,\Theta\}\approx\{2.5,0.8\}$, and $\delta n\approx 0.06$ for $\{q,\Theta\}\approx\{2.75,0.5\}$. As an illustration, Fig. 3 shows the values of $\delta n(q,\Theta)$ found for the points $\{q,\Theta\}=\{1,1\}$ (unfilled symbols) and $\{q,\Theta\}=\{2.5,0.8\}$ (filled symbols). The figure demonstrates the fair agreement between the theoretical estimate (8) and the Monte Carlo calculations.

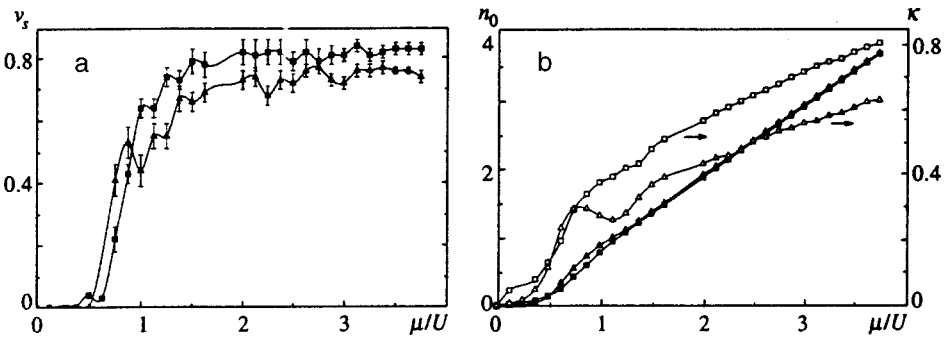


FIG. 4. (a) The superfluid density ν_s as a function of the chemical potential μ/U. (b) Average particle density n_0 (filled symbols) and compressibility modulus κ (unfilled symbols) as functions of the chemical potential μ/U: squares—$\{U/t,k_BT/t\}=\{2.0,0.8\}$, triangles—$\{U/t,k_BT/t\}=\{2.5,0.8\}$. Spline interpolations are shown for visual convenience. The statistical errors are not indicated, but are smaller than the sizes of the corresponding symbols.

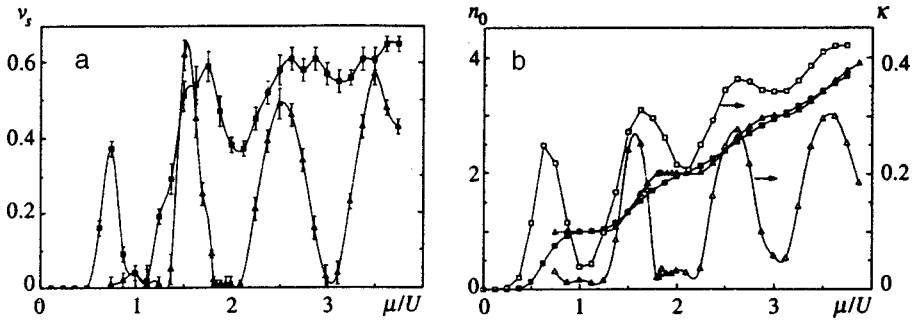


FIG. 5. (a) The superfluid density ν_s as a function of the chemical potential μ/U. (b) Average particle density n_0 (filled symbols) and compressibility modulus κ (unfilled symbols) as functions of the chemical potential μ/U: squares—$\{U/t, k_BT/t\}=\{3.5,0.8\}$, triangles —$\{U/t, k_BT/t\}=\{5.0,0.8\}$.

The above results yield the phase diagram of the system shown in Fig. 1b (in the coordinates $\{\sqrt{U/t}, \mu/T\}$ for $k_BT/t=0.8$) and in Fig. 2b (in the coordinates $\{q,\Theta\}$). The location of the boundary of the ordered superconducting state was estimated from the universal jump in the superfluid density[30] and from the location of the peak in its temperature derivative.[35]

Quantum phase transitions (at $T=0$) in the two-dimensional Hubbard model (1) are determined by the critical properties of the corresponding effective three-dimensional system.[17,32] At finite temperatures a Josephson array will display Kosterlitz–Thouless critical behavior on some $\Theta_c(\mu/U,q)$ curve. We now try to evaluate the effect of quantum fluctuations on the temperature of this transition, assuming it to be rather low (see below). There is interest in two cases: (a) the system has a commensurate particle density $n_0=k$ and at some temperature $\Theta_c(q)$ it undergoes a transition from the superconducting to the normal state, and (b) a phase transition takes place at $\Theta_c(n_0)$ because of a change in the density n_0.

For the region near the point $q_c^{XY}\approx 2.5$ of the quantum transition in the $2+1$-dimensional XY model (see Fig. 2b), the temperature $\Theta_c(q)$ of the topological Kosterlitz–Thouless phase transition has been estimated[24] as $\Theta_c(q)\sim|q-q_c^{XY}|^{\zeta_1}$ with $\zeta_1\approx 0.67$. This estimate was derived under the assumption that the system temperature is less than the temperature of the $2D\rightarrow 3D$ crossover; i.e., $\Theta\leqslant\Theta_{3D}$, where $\Theta_{3D}\sim|q-q_c^{XY}|^{\nu}$ with $\nu\approx\zeta_1$. This makes it possible to estimate the density of the superfluid component, $\nu_s(q,\Theta)$, which determines the phase-transition temperature $\Theta_c(q)$, as $\nu_s(q,\Theta)\approx\nu_s(q,0)$. Evidently, similar arguments apply in the neighborhood of the points $q_c^H|_{n_0=k}$ of the quantum phase transitions in the Hubbard model (1) for integer populating of the array granules, i.e., $n_0=k$ ($q_c^H|_{n_0=1}\approx 2.8$; see Ref. 20). Thus, we may expect that

$$\Theta_c(q;k)\sim|q_c^H|_{n_0=k}-q|^{0.67}. \tag{9}$$

Similar arguments can also be applied to the case of incommensurate boson densities, $n_0\neq k$. It has been shown[17,20] that $\nu_s\sim|n_0-k|^{\zeta_2}$ with $\zeta_2\approx 1.0$ for $q>q_c^H$. Thus, the following relation holds for the temperature $\Theta_c(n_0)$ of the topological Kosterlitz–Thouless phase transition:

$$\Theta_c(n_0)\sim|n-k|^{1.0}. \tag{10}$$

Our quantum calculations are in qualitative agreement with the predictions of Eqs. (9) and (10), but it is difficult to confirm their validity with sufficient accuracy because of the large errors in determining the position of the quantum phase-transition line $\mu(U/t;T=0)$.

There is great interest in the question of the existence of reentrant superconductivity, for which, within some range of variation of the quantum parameter q, disorder sets in not only as the temperature Θ is lowered, but also as it is raised. The existence of reentrant effects has been predicted a number of times within the quantum XY model (see Ref. 13 and the literature cited therein), but as far as we know, computer simulation cannot unequivocally confirm[31,36] or refute[34] the existence of this phenomenon. The earlier numerical calculations of the Hubbard model did not reveal low-temperature

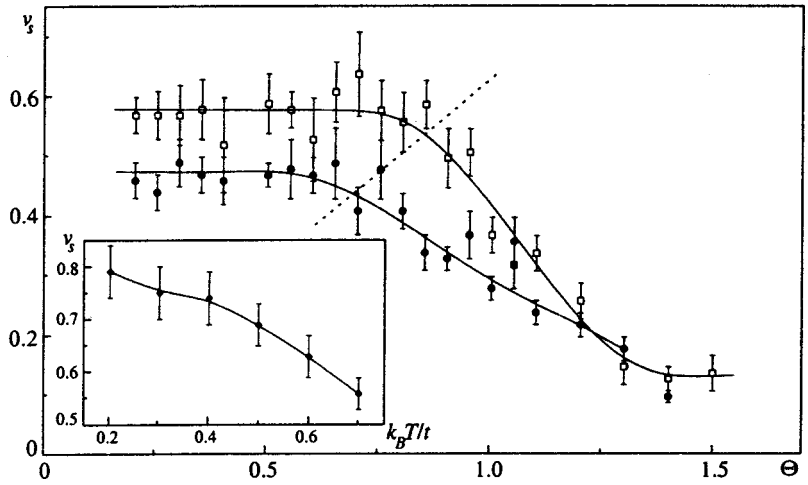


FIG. 6. Superfluid density of the system as a function of $\Theta=k_BT/tn_0$ for $q=3.0$; unfilled symbols—$n_0=1.306$; filled symbols—$n_0=1.194$. The dotted line indicates the straight line $\nu_s=2\Theta/\pi$. The inset shows the calculated superfluid density of the system as a function of k_BT/t for $\sqrt{U/t}=2.5$ and $\mu/U=0.75$.

instability or reentrant superconductivity.[20,23] For this paper we studied the low temperature region $q \approx 2.5$, $\Theta < 0.5$ of the boson Hubbard model both for different noninteger occupation numbers and for fixed values of the chemical potential. The results, shown in Fig. 6, indicate a lack of reentrant superconductivity effects, at least within the range of variation of the control parameters studied here. Special attention was devoted to the region $\{\sqrt{U/t}, \mu/U\} \approx \{2.5, 0.8\}$, within which low-temperature disorder has been predicted.[19] The numerical simulations clearly showed that there were no such effects within this region. In addition, our calculation of the boundary of the ordered state using mean-field theory (see Sec. 2) also disagrees with these predictions. Note that Eq. (2), which specifies the boundary of the ordered state, is more accurate in this method than is the equation used in the paper just cited inasmuch as its derivation did not rely on the assumption that $n_0 \gg 1$ [the latter is valid for $\mu/U \gg 1$ (Ref. 19)].

We conclude by discussing the basic results of this paper. We have analyzed the effect of quantum fluctuations in the phase and modulus of the superconducting or superfluid order parameter on the character of the ordering in two-dimensional mesoscopic Josephson and granular systems within a lattice boson Hubbard model. Quantum Monte Carlo calculations have been used to show that the way the system properties change as a result of modulation of the average occupation number of the array elements by the chemical potential (the substrate potential) is determined by the parameter $q = \sqrt{U/tn_0}$ (i.e., the ratio of the characteristic Coulomb energy of a granule to the Josephson tunneling energy). For $q < 1.5$, which is the quasiclassical region for the quantum XY model and the region of strong fluctuations in the modulus of the order parameter for the Hubbard model (1), the system properties are insensitive to the average number of particles in the granules. In the region where there are significant quantum fluctuations in the order parameter ($q > 2$, $\Theta < 0.8$), we have found that the state of the system depends (more distinctly at lower temperatures) on the average number of particles in it.

This work was supported by grants from the Russian Fund for Fundamental Research and the Program on the Physics of Solid-State Nanostructures.

*)E-mail: lozovik@isan.troitsk.ru

1)Two-dimensional Josephson arrays (of mesoscopic elements) with superfluid helium can, in principle, be formed by creating the corresponding ''picture'' in cesium on the substrate (since cesium is not wetted by helium).[1]

2)The interference of two Bose condensates has recently been studied.[11]

[1] F. J. Nacker and J. Dupont-Roc, Phys. Rev. Lett. **67**, 2966 (1991).
[2] J. D. Reppy, J. Low Temp. Phys. **67**, 207 (1992).
[3] H. S. J. van der Zant, F. C. Fritschy, J. E. Mooij *et al.*, Phys. Rev. Lett. **69**, 2971 (1992); J. E. Mooij, R. Fazio, G. Schön *et al.*, Phys. Rev. Lett. **65**, 645 (1990).
[4] V. G. Gantmakher, V. M. Teplinskiĭ, and V. N. Zverev, JETP Lett. **62**, 887 (1995).
[5] A. F. Hebard and M. A. Paalanen, Phys. Rev. Lett. **65**, 927 (1990).
[6] A. L. Dobryakov, Yu. E. Lozovik, A. A. Puretzky *et al.*, Appl. Phys. A **54**, 100 (1992).
[7] Yu. M. Mucharsky, A. Loshak, K. Schwab *et al.*, Czech. J. Phys. **46**, 115 (1996); S. V. Pereverzev, A. Loshak, S. Backhaus *et al.*, Nature (London) **388**, 449 (1997).
[8] M. N. Anderson, J. R. Ensher, M. R. Mathews *et al.*, Science **269**, 198 (1995).
[9] C. C. Bradley, C. A. Sackoff, J. J. Tollett *et al.*, Phys. Rev. Lett. **75**, 1687 (1995).
[10] K. B. Davis, M.-O. Mewes, M. R. Andrew *et al.*, Phys. Rev. Lett. **75**, 3969 (1995).
[11] M. R. Andrews, C. G. Towsend, J.-J. Miesner *et al.*, Science **275**, 637 (1997).
[12] Yu. E. Lozovik, submitted to Physica E (Amsterdam); Yu. E. Lozovik and O. L. Berman, Zh. Éksp. Teor. Fiz. **111**, 1879 (1997) [JETP **84**, 1027 (1997)]; Yu. E. Lozovik, O. L. Berman, and V. G. Tsvetus, JETP Lett. **66**, 355 (1997).
[13] B. J. Kim and M. Y. Choi, Phys. Rev. B **52**, 3624 (1995); B. J. Kim, J. Kim, M. Y. Choi *et al.*, Phys. Rev. B **56**, 395 (1997).
[14] C. Bruder, R. Fazio, A. P. Kampf *et al.*, Phys. Scr. T **42**, 159 (1992).
[15] I. E. Dzyaloshinskii, E. M. Lifshitz, and L. P. Pitaevskii, Adv. Phys. **10**, 165 (1961).
[16] G. T. Zimanyi, P. A. Crowell, R. T. Scalettar *et al.*, Phys. Rev. B **50**, 6515 (1994).
[17] M. P. A. Fisher and G. Grinstein, Phys. Rev. Lett. **60**, 208 (1988); M. P. A. Fisher, P. B. Weichman, G. Grinstein, and D. S. Fisher, Phys. Rev. B **40**, 546 (1989); M. P. A. Fisher, G. Grinstein, and S. M. Girvin, Phys. Rev. Lett. **64**, 587 (1990).
[18] M. C. Cha, M. P. A. Fisher, S. M. Girvin *et al.*, Phys. Rev. B **44**, 6883 (1991).
[19] A. P. Kampf and G. T. Zimanyi, Phys. Rev. B **47**, 279 (1993).
[20] W. Knauth, N. Trivedi, and D. Ceperley, Phys. Rev. Lett. **67**, 2703 (1991); W. Krauth and N. Trivedi, Europhys. Lett. **14**, 627 (1991).
[21] V. A. Kashurnikov, A. V. Krasavin, and B. V. Svistunov, JETP Lett. **64**, 99 (1996).
[22] A. V. Otterlo and K. H. Wagenblast, Phys. Rev. Lett. **72**, 3598 (1994); E. Roddick and D. Stroud, Phys. Rev. B **51**, 8672 (1995).
[23] A. I. Belousov, S. A. Verzakov, and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **113**, 261 (1998) [JETP **86**, 146 (1998)]; A. I. Belousov and Yu. E. Lozovik, JETP Lett. **66**, 686 (1997).
[24] S. Doniach, Phys. Rev. B **24**, 5063 (1981).
[25] V. N. Popov, *Functional Integrals in Quantum Field Theory and Statistical Physics*, Reidel, Dordrecht (1983).
[26] J. J. Alvarez and C. A. Balseiro, Solid State Commun. **98**, 313 (1996).
[27] P. Olsson, Phys. Rev. B **52**, 4511 (1995).
[28] A. Blaer and J. Han, Phys. Rev. A **46**, 3225 (1992).
[29] G. G. Batrouni, B. Larson, R. T. Scalettar *et al.*, Phys. Rev. B **48**, 9628 (1993).
[30] P. Minnhagen, Rev. Mod. Phys. **59**, 1001 (1987).
[31] M. Jacobs, J. V. Jose, M. A. Novotny *et al.*, Phys. Rev. B **38**, 4562 (1988).
[32] S. L. Sondhi, S. M. Girvin, J. P. Carini *et al.*, Rev. Mod. Phys. **69**, 315 (1997).
[33] A. I. Belousov and S. G. Akopov, J. Phys. C **14**, L31 (1981); S. G. Akopov and Yu. E. Lozovik, J. Phys. C **15**, 4403 (1982).
[34] A. I. Belousov and Yu. E. Lozovik, Solid State Commun. **100**, 421 (1996); A. I. Belousov and Yu. E. Lozovik, Fiz. Tverd. Tela (St. Petersberg) **39**, 1513 (1997) [Phys. Solid State **39**, 1345 (1997)]; S. A. Verzakov and Yu. E. Lozovik, Fiz. Tverd. Tela (St. Petersberg) **39**, 818 (1997) [Phys. Solid State **39**, 724 (1997)].
[35] F. F. Assaad, W. Hanke, and D. J. Scalapino, Phys. Rev. B **50**, 12 835 (1994).
[36] D. Marx and P. Nielaba, J. Chem. Phys. **102**, 4538 (1995).

Translated by D. H. McNeill
Edited by P. Shelnitz

Theory of the pseudogap formation in 2D attracting fermion systems

V. M. Loktev*)

Bogolyubov Institute for Theoretical Physics, 252143 Kiev, Ukraine

V. M. Turkowski

Shevchenko Kyiv University, 252127 Kiev, Ukraine

(Submitted 1 August 1997)
Zh. Éksp. Teor. Fiz. **114**, 605–618 (August 1998)

The two-dimensional fermion system with the indirect Einstein phonon-exchange attraction and additional local four-fermion interaction is considered. It is shown that as a result of the attraction between fermions, the normal phase of such a system is divided into two regions. In one of them, called the pseudogap region, the absolute value of the order parameter exists as essentially nonzero value, but its phase is a random quantity. It is important that in the case of attraction due to the phonons, this abnormal region appears at rather low carrier concentrations, i.e., it decreases appreciably with increasing doping. The relevance of the results obtained for high-temperature superconductors is speculated. © *1998 American Institute of Physics.* [S1063-7761(98)01708-9]

1. INTRODUCTION

An adequate description of the physical properties of high-temperature superconductors (HTSCs) still remains one of the important problems of modern solid-state physics. It is connected with some peculiar properties of HTSCs. Among them there are such problems as quasi-2D character of electronic (and magnetic) properties, a relatively low and changeable carrier density n_f, and its influence on the properties of HTSCs (see, for example, the review article[1]).

At present, one of the widely discussed topics on HTSCs is the ''pseudogap'' (or ''spin gap'' if magnetic subsystem of HTSCs is taken into account),[2–4] which is experimentally observed, for example, as a loss in the spectral weight of quasiparticle (or spin) excitations in the normal-state samples with lowered carrier density.[5–7] Corresponding samples reveal some specific spectral, magnetic, and thermodynamic peculiarities which are not yet sufficiently understood. In addition, the striking difference between the low (underdoped) and high (overdoped) density regions in HTSCs is hotly debated and is considered as one of the central and key topics considered in the physics of cuprates.[8,9]

The possibility of experimentally changing the carrier concentration in HTSCs created a general theoretical problem of the description of the crossover from composite boson superfluidity (low n_f) to Cooper pairing (large n_f) when n_f increases (in other words, a description of the continuous transition from the so-called underdoped regime to the overdoped one). Such a crossover has already been studied in 3D and quasi-2D systems (see the review articles[10,11]). The 2D case has been considered[10,12] at temperature $T=0$ in connection with the Honenberg–Mermin–Wagner theorem, which forbids any homogeneous, i.e., long-range) order in pure 2D systems at $T\neq 0$ due to the long-wave fluctuations of the charged order parameter (OP).

The problem of the inhomogeneous condensate [the Berezinskii–Kosterlitz–Thouless (BKT) phase] formation was also considered, despite some difficulties in the $2+1$ relativistic field models,[13] where the fermion concentration effects are irrelevant. At the same time, these effects were studied in the nonrelativistic model in Ref. 14, for example, without allowance for the existence of the neutral order parameter ρ. Its consideration proves to be very important (see Ref. 15) and, in fact, results in the formation of an equilibrium region with $\rho\neq 0$, which is mainly located in the phase diagram of a system between the ordinary normal phase and the superconducting (here BKT) phase. Because of the fluctuations of the OP phase, this new region of the system, which is a part of the normal phase, is of course a nonsuperconducting phase.

In this paper we attempt to study the crossover and the possibility for the appearance of the above-mentioned new region in the 2D fermion system with a more realistic indirect (phonon) and also a direct (local) four-fermion (4F) interactions. Thus, this study is to a certain extent a specific and nontrivial generalization of the preliminary short communication,[15] where this abnormal region was studied for the 4F case only, and of the paper,[16] in which the Fröhlich model was used to study the crossover at $T=0$. As will be seen in the boson-exchange model (in contrast with the pure 4F case), the new region exists when n_f is rather small, which allows one to compare this result qualitatively with the underdoped HTSC compounds. It is actually more interesting to take into account a more realistic situation with an indirect attraction and some kind of local repulsion, which may in principle correspond to the short-range (screened) Coulomb interaction between carriers. In general, however, we assume that 4F interaction can be repulsive as well as attractive. In addition, the case of total repulsion allows one to explore the fermion-antifermion (electron-hole) pairing channel, which, despite a physical difference, can be formally described in the same manner.

2. MODEL AND BASIC EQUATIONS

Let us choose the simplest Hamiltonian density in the form

$$H(x)=-\psi_\sigma^\dagger(x)\left(\frac{\nabla^2}{2m}+\mu\right)\psi_\sigma(x)H_{\rm ph}(\varphi(x))+g_{\rm ph}\psi_\sigma^\dagger(x)\psi_\sigma(x)\varphi(x)-g_{4\rm F}\psi_\uparrow^\dagger(x)\psi_\downarrow^\dagger(x)\psi_\downarrow(x)\psi_\uparrow(x),\quad (x=\mathbf{r},t),\qquad (1)$$

where $\psi_\sigma(x)$ is the fermionic field with an effective mass m and spin $\sigma=\uparrow,\downarrow$, μ is the chemical potential of the fermions which fixes n_f, $\varphi(x)$ is a phonon field operator, and $g_{\rm ph}$ and $g_{4\rm F}$ are the electron-phonon and the 4F interaction coupling constants, respectively. As was indicated above, $g_{4\rm F}$ can be positive (fermion-fermion attraction) or negative (fermion-antifermion attraction); in Eq. (1) we set $\hbar=k_B=1$.

In Eq. (1) $H_{\rm ph}$ is the Hamiltonian of free phonons, which can be described by the propagator

$$D(i\Omega_n)=-\frac{\omega_0^2}{\Omega_n^2+\omega_0^2},\qquad (2)$$

where $\Omega_n=2n\pi T$ (n is an integer) is the Matsubara frequency.[17] As follows from (2), the propagator $D(i\Omega_n)$ was chosen in a very simple form; here ω_0 is the Einstein (dispersionless) phonon frequency. This choice was made for several reasons. First, this propagator makes it possible to integrate the equations which we obtained. Second, the optic phonon and quadrupolar exciton modes with their relatively weak dispersion are widely considered as exchange bosons which can contribute to the hole-hole attraction in HTSCs.[1,18,19] Third, the qualitative results concerning retardation effects do not strongly depend on the model studied. On the other hand, the propagator (2) for the model under consideration can hardly be used for quantitative description of the cuprates and their spin-wave branches, which, as is well known, obey the linear dispersion relation.

It is important that the Hamiltonian (1) is invariant under global gauge transformations of two types:[20]

$$\psi_\sigma(x)\to\psi_\sigma(x)e^{i\alpha},\quad \psi_\sigma^\dagger(x)\to\psi_\sigma^\dagger(x)e^{-i\alpha},\qquad (3)$$

and

$$\psi_\uparrow(x)\to\psi_\uparrow(x)e^{i\alpha},\quad \psi_\downarrow(x)\to\psi_\downarrow(x)e^{-i\alpha},$$
$$\psi_\uparrow^\dagger(x)\to\psi_\uparrow^\dagger(x)e^{-i\alpha},\quad \psi_\downarrow^\dagger(x)\to\psi_\downarrow^\dagger(x)e^{i\alpha},\qquad (4)$$

which must be taken into account. The phase α in (3) and (4) is real.

To calculate the phase diagram of a system it is necessary to find its thermodynamic potential. It can be calculated by making use of the auxiliary bilocal field method (see, for example, Ref. 21), which is a generalization of the standard Hubbard-Stratonovich method for the boson-exchange case. The grand partition function Z can then be expressed in terms of the path integral over the fermionic $\psi_\sigma(x)$ and the complex auxiliary fields [for example, $\phi(x,x')\sim\langle\psi_\uparrow^\dagger(x)\psi_\downarrow^\dagger(x')\rangle$].

In the case of model (1) it is convenient, following Ref. 22, to introduce the bispinor

$$\Psi^\dagger(x)=(\psi_\uparrow^\dagger(x),\psi_\downarrow^\dagger(x),\psi_\uparrow(x),\psi_\downarrow(x))\qquad (5)$$

and its Hermitian conjugate, which here are the analogs of the Nambu spinors.[23] After substituting (5) in (1), we can write the Hamiltonian in the form

$$H(x)=-\frac{1}{2}\Psi^\dagger(x)\left(\frac{\nabla^2}{2m}+\mu\right)I\otimes\tau_z\Psi(x)-\frac{1}{2}g_{\rm ph}\Psi^\dagger(x)I\otimes\tau_z\Psi(x)\varphi(x)-\frac{1}{4}g_{4\rm F}\Psi^\dagger(x)I\otimes\tau_z\Psi(x)\Psi^\dagger(x)I\otimes\tau_z\Psi(x)+\varphi(x)D^{-1}(x)\varphi(x),\qquad (6)$$

where $I\otimes\tau_z$ is the direct product of the unit I and Pauli τ_z 2×2 matrices, and $D(x)$ is defined by (2). In such a representation of the Hamiltonian (6) and the field variables (5) the Feynman diagram technique is applicable in the usual form.[22] Thus, after standard exclusion of the boson field $\varphi(x)$, the Lagrangian of the system can be expressed by the formula

$$L(x_1,y_1,x_2,y_2)=\frac{1}{2}\Psi^\dagger(x)\left[-\partial_\tau+\left(\frac{\nabla^2}{2m}+\mu\right)I\otimes\tau_z\right]\times\Psi(x)-\frac{1}{4}\Psi(x_1)\Psi^\dagger(y_1)I\otimes\tau_zK(x_1,y_1;x_2,y_2)\Psi(x_2)\times\Psi^\dagger(y_2)I\otimes\tau_z.\qquad (7)$$

The kernel K is the effective, nonlocal, particle-particle interaction function which is explicitly defined in the momentum space below.

In order to explore the pairing possibility in the system we introduce the bilocal auxiliary field or OP.

$$\phi(x_1,y_1)=\tau_zK(x_1,y_1;x_2,y_2)\Psi(x_2)\Psi^\dagger(y_2)I\otimes\tau_z\equiv-i\tau_+\otimes\tau_y\phi_{\rm ch}^*(x_1,y_1)-i\tau_-\otimes\tau_y\phi_{\rm ch}(x_1,y_1)-\tau_z\otimes I\phi_{\rm ins}(x_1,y_1),\qquad (8)$$

where $\tau_+=(\tau_x+i\tau_y)/2$, $\tau_-=(\tau_x-i\tau_y)/2$, and the integration over x_2 and y_2 is assumed. Here $\phi_{\rm ch}\sim\langle\psi_\downarrow^\dagger\psi_\uparrow^\dagger\rangle$ and $\phi_{\rm ins}\sim\langle\psi_\uparrow^\dagger\psi_\uparrow\rangle$ are the electron-electron (charged) and the electron-hole (insulating) spin-singlet OP, respectively (we ignore the nonzero spin pairing). The auxiliary fields $\phi_{\rm ch}$ and $\phi_{\rm ins}$ are responsible for the dynamic symmetry breaking [in accordance with (3) and (4), respectively].

Adding to (7) a zero term

$$\frac{1}{4}[\phi(x_1,y_1)-K(x_1,y_1;x_1',y_1')\Psi(x_1')\Psi^\dagger(y_1')I\otimes\tau_z]\times K^{-1}(x_1,y_1;x_2,y_2)[\phi(x_2,y_2)-K(x_2,y_2;x_2',y_2')\Psi(x_2')\Psi^\dagger(y_2')I\otimes\tau_z]$$

in order to cancel the 4F interaction, we obtain the Lagrangian in the form

$$L(x_1,y_1;x_2,y_2)$$
$$=\frac{1}{2}\Psi^{\dagger}(x_1)\Big[-\partial_\tau+\Big(\frac{\nabla^2}{2m}+\mu\Big)I\otimes\tau_z$$
$$-\frac{1}{2}I\otimes\tau_z\phi(x_1,y_1)\Big]\Psi(y_1)+\frac{1}{4}\,\phi(x_1,y_1)K^{-1}$$
$$\times(x_1,y_1;x_2,y_2)\,\phi(x_2,y_2). \tag{9}$$

Let us transform the expression for the kernel K; in the momentum space it then is

$$K(x_1,y_1;x_2,y_2)=\int\frac{d^3Pd^3p_1d^3p_2}{(2\pi)^9}K_P(p_1;p_2)$$
$$\times\exp\Big[-iP\Big(\frac{x_1+y_1}{2}-\frac{x_2+y_2}{2}\Big)$$
$$-ip_1(x_1-y_1)-ip_2(x_2-y_2)\Big],$$

where $p_i=(\mathbf{p}_i,\omega_i)$ $(i=1,2)$ and $P=(\mathbf{P},\omega)$ represent the relative and the center-of-mass momenta, respectively. According to the definition, the kernel $K_P(p_1;p_2)$ is in fact independent of P (we can therefore omit the index P below) and acquires a simple form

$$K(p_1;p_2)=g_{\rm ph}^2D(p_1-p_2)-g_{4\rm F}, \tag{10}$$

which is used in (9). The last expression evidently demonstrates that the total character of the effective particle-particle interaction, as it always takes place in such a situation,[23,24] is defined by a possible competition between the first (retarded) and the second (nonretarded) terms in (10) or, in other words, by their common action.

The partition function can be written as

$$Z=\int\mathscr{D}\Psi^{\dagger}\mathscr{D}\Psi\mathscr{D}\phi\mathscr{D}\phi^*$$
$$\times\exp\Big[-\beta\int L(\Psi^{\dagger},\Psi,\phi^*,\phi)dxdy\Big]$$
$$\equiv\int\mathscr{D}\phi\mathscr{D}\phi^*\,\exp(-\beta\Omega[\mathscr{G}]),\quad\beta=1/T,$$

where $\Omega[\mathscr{G}]$ is the thermodynamic potential which in the ''leading order'' is

$$\beta\Omega[\mathscr{G}]=-\mathrm{Tr}\Big[\ln\mathscr{G}^{-1}+\frac{1}{2}\,(\phi K^{-1}\phi)\Big], \tag{11}$$

where Tr includes 2D spatial $\mathbf{r}$ and ''time'' $0\leqslant\tau\leqslant\beta$ integrations, as well as the standard trace operation. The complete Green's function of the system is

$$\mathscr{G}^{-1}=-\frac{1}{2}\Big[\partial_\tau-\Big(\frac{\nabla^2}{2m}+\mu\Big)I\otimes\tau_z-\phi\Big]. \tag{12}$$

From (11) and (12) we obtain the ϕ-equation (the Schwinger–Dyson equation)

$$\frac{\delta\Omega}{\delta\phi}=\phi-\int\frac{d^2\mathbf{k}d\omega}{(2\pi)^3}K(p;\mathbf{k},\omega)\mathscr{G}(\mathbf{k},\omega)=0. \tag{13}$$

Substituting (13) into (11), we obtain the expression for $\Omega(\mathscr{G})$

$$\beta\Omega(\mathscr{G})=-\mathrm{Tr}\,\ln\mathscr{G}^{-1}+\frac{1}{2}\,\mathrm{Tr}\,\mathscr{G}K\mathscr{G}.$$

This expression is the standard Cornwell–Jackiw–Tomboulis formula for the effective action in the one-loop approximation.[25] Using (13), we can rewrite this expression in the form

$$\beta\Omega(\mathscr{G})=-\mathrm{Tr}\Big[\ln\mathscr{G}+\frac{1}{2}\,[\mathscr{G}\mathscr{G}_0^{-1}-1]\Big]. \tag{14}$$

As was shown by Thouless *et al.*[26] (see also Ref. 15) in the 2D case it is more logical to use a new parametrization of the charge OP [Eq. (8)]—its absolute value (modulus) and the phase. In other words,[1)]

$$\phi_{\rm ch}(x,y)=\rho_{\rm ch}(x,y)\exp[-i(\theta(x)+\theta(y))/2], \tag{15}$$

where $\rho_{\rm ch}$ is real. As for $\rho_{\rm ins}$, it corresponds, as can be seen from Eq. (8), to a one-component OP and therefore does not characterize the phase factor.

As will be shown below, with the given kernel (10) only one ($\phi_{\rm ch}$ or $\phi_{\rm ins}$) OP can arise. Therefore, it is necessary to make, simultaneously with (15), the spinor transformation [in accordance with (3) and (4)]

$$\Psi^{\dagger}(x)=\chi^{\dagger}(x)\exp(i\theta(x)I\otimes\tau_z/2), \tag{16}$$

$$\Psi^{\dagger}(x)=\chi^{\dagger}(x)\exp(i\theta(x)\tau_z\otimes\tau_z/2) \tag{17}$$

[the spinor $\chi(x)$ is real and formally corresponds to chargeless fermions]. Below we shall obtain the θ-corrections for the $\phi_{\rm ch}$ case only, but the equations for $\rho_{\rm ins}$ are the same up to the substitution $\rho_{\rm ch}\to\rho_{\rm ins}$. The reason is that when $K(p_1,p_2)$ describes the attraction (charge pairing channel), the symmetry of the Lagrangian under operations (3) proves to be crucial for the representation (16); but when $K(p_1,p_2)$ corresponds to the repulsion (chargeless or electron-hole pairing channel), the symmetry (4) is already important and the representation (17) must be used as a ''working'' representation. With this difference, the rest of the calculations for ρ's are almost identical but the ''phase effects'' persist for the charge channel only. We shall therefore examine in detail this channel, the most interesting one for metallic (superconducting) systems.

In the variables (16) the Green's function (12) transforms to

$$\mathscr{G}^{-1}=-\frac{1}{2}\Big[\partial_\tau-I\otimes\tau_z\Big(\frac{\nabla^2}{2m}+\mu\Big)+i\tau_x\otimes\tau_y\rho_{\rm ch}-I$$
$$\otimes\tau_z\Big(\partial_\tau\theta+\frac{\nabla\theta^2}{2m}\Big)-iI\otimes I\Big(\frac{\nabla^2\theta}{2m}+\frac{\nabla\theta\nabla}{m}\Big)\Big]$$
$$\equiv G^{-1}(\rho_{\rm ch})-\Sigma(\partial\theta). \tag{18}$$

Using (18) under assumption that the θ gradients are small (the hydrodynamic approximation) and taking them into account up to the second order, we can divide the effective potential (14) into two parts: $\Omega=\Omega_{\rm kin}(\rho_{\rm ch},\nabla\theta)+\Omega_{\rm pot}(\rho_{\rm ch})$, where in the $(\nabla\theta)^2$ approximation

$$\beta\Omega_{\rm kin}(\rho_{\rm ch},\nabla\theta)={\rm Tr}\Big[G\Sigma-G_0\Sigma+\frac{1}{2}\,G\Sigma G\Sigma-\frac{1}{2}\,G_0\Sigma G_0\Sigma + \tau_x\otimes I\,\frac{1}{2}\,i\rho_{\rm ch}G(G\Sigma+G\Sigma G\Sigma)\Big]. \quad (19)$$

Assuming by analogy with Ref. 27 (see also Ref. 28) that $\rho_{\rm ch}(x,y)$ is homogeneous[2)] after rather tedious but otherwise straightforward calculation, we find from (19) the expression

$$\Omega_{\rm kin}(\rho_{\rm ch},\Delta\theta)=\frac{T}{2}\times\int_0^\beta d\tau\int d^2\mathbf{r}J(\mu,T,\rho_{\rm ch}(\mu,T))(\nabla\theta)^2, \quad (20)$$

where

$$J(\mu,T,\rho_{\rm ch}(\mu,T))=\frac{1}{8\pi}\Bigg(\sqrt{\mu^2+\rho_{\rm ch}^2}+\mu+2T\times\ln\Bigg[1+\exp\Bigg(-\frac{\sqrt{\mu^2+\rho_{\rm ch}^2}}{T}\Bigg)\Bigg]-\frac{T}{4\pi}\Bigg[1-\frac{\rho_{\rm ch}^2}{4T^2}\,\frac{\partial}{\partial(\rho_{\rm ch}^2/4T^2)}\Bigg]\times\int_{-\mu/2T}^{\infty}dx\,\frac{x+\mu/2T}{\cosh^2\sqrt{x^2+\rho_{\rm ch}^2/4T^2}}, \quad (21)$$

which plays the role of the neutral OP stiffness. Note that in comparison with the retardation-free 4F model,[15] the last expression contains one more term: the term with the derivative.

The equation for the temperature $T_{\rm BKT}$ of the BKT transition can be written, after direct comparison of the kinetic term (20), in the effective action with the Hamiltonian of the 2D *XY* model, which formally has the identical form.[29] It is therefore easy to conclude that

$$\frac{\pi}{2}\,J(\mu,T_{\rm BKT},\rho_{\rm ch}(\mu,T_{\rm BKT}))=T_{\rm BKT}. \quad (22)$$

The basic difference between this equation and the one for the *XY* model is the inherent dependence of the former on μ (or n_f) and $\rho_{\rm ch}$.

To complete the set of self-consistent equations, which allow one to trace an explicit dependence of $T_{\rm BKT}$ on n_f, we also give the equations for $\rho_{\rm cosh}$ and μ. In particular, a simple equation for $\rho_{\rm ch}(i\omega_n)$ is Eq. (13) with $\nabla\theta=0$; i.e., the Green's function G of the neutral fermions substitutes $\mathscr{G}$, so that (13) in the frequency-momentum representation takes the form

$$\begin{pmatrix}\rho_{\rm ch}(i\omega_n)\\ \rho_{\rm ins}(i\omega_n)\end{pmatrix}=T\sum_{m=-\infty}^{\infty}\int\frac{d^2\mathbf{k}}{(2\pi)^2}\begin{pmatrix}-\rho_{\rm ch}(i\omega_m)\\ +\rho_{\rm ins}(i\omega_m)\end{pmatrix}\times\frac{K(\omega_n,\omega_m)}{\omega_m^2+\xi^2(\mathbf{k})+\rho_{\rm ch}^2(i\omega_m)+\rho_{\rm ins}^2(i\omega_m)}, \quad (23)$$

where $\omega_n=(2n+1)\pi T$ is the Matsubara fermionic frequency,[23] $\xi(\mathbf{k})=\mathbf{k}^2/2m-\mu$, and the kernel $K(\omega_m,\omega_n)$ is defined above.

We gave the final equations for both OPs, $\rho_{\rm ch}$ and $\rho_{\rm ins}$ in order to show that they indeed are the same but alternative if the kernel K changes sign. The analytic solution of these equations, of Eq. (22), and the number equation can be done by assuming that $\rho_{\rm ch}(i\omega_n)$ does not depend on the Matsubara frequencies (see the footnote on p. 6).

Making use of this approximation, the equation which follows from the ordinary condition $V^{-1}\partial\Omega[\mathscr{G}]/\partial\mu=-n_f$ (V is the volume of the system) and which is crucial for the crossover description must be added to Eqs. (22) and (23) for self-consistency. We thus obtain

$$\sqrt{\mu^2+\rho_{\rm ch}^2}+\mu+2T\,\ln\Bigg[1+\exp\Bigg(-\frac{\sqrt{\mu^2+\rho_{\rm ch}^2}}{T}\Bigg)\Bigg]=2\epsilon_{\rm F}, \quad (24)$$

where $\epsilon_{\rm F}=\pi n_f/m$ is the Fermi energy of free 2D fermions with a simple quadratic dispersion relation. Thus, in the case under consideration all unknown quantities, $\rho_{\rm ch}$, μ, and $T_{\rm BKT}$, are the explicit functions of n_f.

3. ANALYSIS OF THE SOLUTIONS

In contrast with the standard (the *T*-independent unit vector) *XY* model, in the superconducting model two characteristic temperatures can be introduced: T_ρ, where formally the complete OP given by (8) arises but its phase is a random quantity,[3)] i.e., $\langle\phi(x,y)\rangle=0$ and another temperature, $T_{\rm BKT}<T_\rho$, where the phase of the OP is ordered, so that $\langle\phi(x,y)\rangle\neq0$. In other words, we define the temperature T_ρ as the temperature of a relatively abrupt change in the neutral OP, which does not break any real symmetry. Therefore, this temperature (in contrast with $T_{\rm BKT}$) is not the phase-transition temperature. Nevertheless, it gives (see Refs. 27 and 28) a convenient scale for the description of the neutral OP temperature behavior. Recall that according to the equations obtained above, both these temperatures directly depend on the carrier density in the system.

The ''critical'' temperature T_ρ can be found, for example, from Eqs. (21)–(24) by setting $\rho_{\rm ch}=0$ (in accordance with the derivation of these equations, it corresponds to the mean-field approximation[4)]). As a result, with a decrease in temperature, a 2D metal (similarly to a 1D metal[27]) passes from the normal phase ($T>T_\rho$) to another phase, where the average homogeneous (charged) OP $\langle\phi(x,y)\rangle=0$ or, equivalently, the superconductivity is absent, but chargeless OP $\rho_{\rm ch}\neq0$. It is evident that the pseudogap is formed just in the temperature region $T_{\rm BKT}<T<T_\rho$, because, as follows from the formulas cited above [see, e.g., Eqs. (21)–(24)], $\rho_{\rm ch}=\rho_{\rm ch}(T)$ acquires all the spectral characteristics of a 2D metal in the same way as the superconducting gap $\Delta(T)$ enters into corresponding expressions for ordinary superconductors. It justifies why this region can be called ''the pseudogap phase.'' The density of states near ϵ_F in the pseudogap phase is definitely lower than that in the region of the normal phase with $\rho_{\rm ch}=0$, but does not equal zero as in the superconducting phase. The latter must be checked by

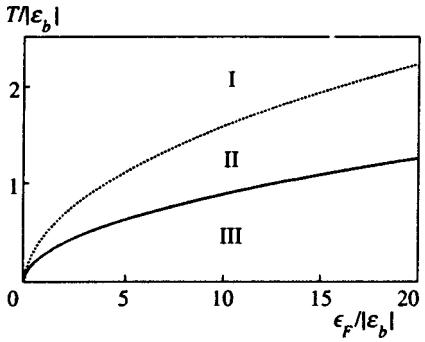


FIG. 1. The $T-n_f$ phase diagram of the 2D metal with 4F fermion attraction. The lines correspond to the functions $T_\rho(n_f)$ (the dotted curve) and $T_{\mathrm{BKT}(n_f)}$ (the solid curve) at $g_{4\mathrm{F}}m/2\pi=0.5$. Figures I, II, and III show the regions of the normal, pseudogap, and superconducting phases, respectively.

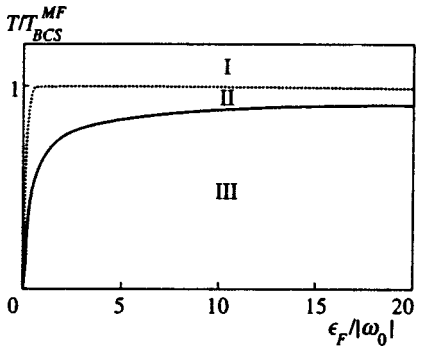


FIG. 2. The $T-n_f$ phase diagram of the 2D metal with indirect intercarrier attraction for $\lambda=0.5$. Similarly to Fig. 1, the curves correspond to the functions $T_\rho(n_f)$ and $T_{\mathrm{BKT}_{(n_f)}}$ and separate the same regions.

direct calculation of the one-particle fermion Green's function, which is most likely a separate problem that is not considered here.

The phase diagram of a system can be found from Eqs. (21)–(24). The quantities $T_\rho(n_f)$ and $T_{\mathrm{BKT}}(n_f)$ behave differently for different correlations between interaction constants.

1) $g_{4\mathrm{F}}>0$, $g_{\mathrm{ph}}=0$ (an unretarded interaction).

This case has been analyzed in part in Ref. 15. It corresponds to fermion-fermion pairing due to the local attraction. Note [see Eq. (23)] that in this case (or in the case of attraction between carriers) the fermion-antifermion (insulating) pairing channel is absent, i.e., $\rho_{\mathrm{ins}}=0$. The corresponding phase diagram is shown in Fig. 1. We see that the pseudogap phase exists at any carrier concentrations, that the temperature width of this phase region increases only slightly with increasing n_f, and that the BKT phase always begins to form when $\rho_{\mathrm{ch}}(T_{\mathrm{BKT}})$ is finite. This means that fluctuations of the latter near and below T_{BKT} are not essential.

As $\epsilon_F\to 0$, the temperature of the BKT phase formation is defined by the equality $T_{\mathrm{BKT}}=\epsilon_{\mathrm{F}}/8$, and T_ρ as a function of n_f can be found from the equation

$$T_\rho \ln(T_\rho/\epsilon_{\mathrm{F}})=W\ \exp(-4\pi/g_{4\mathrm{F}}m)=-\varepsilon_b^{(4\mathrm{F})}/2,$$

which follows from (23) (W is the conduction band width, and $\varepsilon_b^{(4\mathrm{F})}$ is the two-fermion bound-state energy, which is always different from zero in the 2D case).

2) $g_{4\mathrm{F}}=0$, $g_{\mathrm{ph}}\neq 0$ (a pure indirect interaction).

This is one of the most interesting cases because it corresponds to the widely accepted electron-phonon (or the BCS-Bogolyubov–Eliashberg) model of superconductivity. The numerically calculated phase diagram is shown in Fig. 2. It shows that a comparatively large region with the pseudogap phase exists at rather low carrier concentrations only, and that its temperature area shrinks when $n_f\to\infty$. Such a behavior qualitatively agrees with that which takes place in real HTSCs[5–8] and demonstrates that a pseudogap (and also a spin gap) is mainly observed in underdoped HTSC samples.

It is not difficult to conclude that the asymptotic behavior of $T_\rho(n_f)$ and $T_{\mathrm{BKT}}(n_f)$ has the following forms:

i) when the ratio $\epsilon_F/\omega_0\ll 1$ (very low free fermion density or the local pair case), the first one satisfies the equation

$$T_\rho \ln(T_\rho/\epsilon_F)=\omega_0\ \exp(-4\pi/g_{\mathrm{ph}}^2 m)=-\varepsilon_b^{(\mathrm{ph})}/2,$$

which immediately results in $\partial T_\rho(n_f)/\partial n_f|_{n_f\to 0}\to\infty$ (here similarly to the 4F case it is convenient to introduce the bound-state energy $\varepsilon_b^{(\mathrm{ph})}$ for the phonon attraction). At the same time, the temperature T_{BKT} in the limit $n_f\to 0$ has identical dependence on the carrier density and, as above, $T_{\mathrm{BKT}}=\epsilon_F/8$. This simply means that here again it is proportional to the number $n_f/2$ of composite bosons; in this density region $T_\rho/T_{\mathrm{BKT}}\gg 1$ (this inequality is also satisfied for the pure 4F interaction).

ii) in the opposite case $\epsilon_F/\omega_0\gg 1$ (very large fermion density or the Cooper pair case) we easily obtain the standard BCS value:

$$T_\rho=(2\gamma\omega_0/\pi)\exp(-2\pi/g_{\mathrm{ph}}^2 m)\equiv T_{\mathrm{BCS}}^{\mathrm{MF}}=(2\gamma/\pi)\Delta_{\mathrm{BCS}}$$

(Δ_{BCS} is the usual one-particle BCS gap at $T=0$). In other words, in this limit the temperature T_ρ is equal, as it should be, to the BCS value.[5)] The T_{BKT} asymptotic behavior here is not so evident and requires a more detailed consideration.

First of all, it is natural to assume that for large n_f the temperature $T_{\mathrm{BKT}}\to T_\rho$. It is then necessary to check the dependence of ρ on T as $T\to T_\rho$. For this purpose Eq. (23) can be transformed to

$$\frac{2\pi}{g_{\mathrm{ph}}^2 m}=\int_0^\infty dx\left(\frac{\tanh\sqrt{x^2+\rho_{\mathrm{ch}}^2/4T^2}}{\sqrt{x^2+\rho_{\mathrm{ch}}^2/4T^2}}-\frac{\tanh\sqrt{x^2+\rho_{\mathrm{ch}}^2/4T^2}-\tanh(\omega_0/2T)}{2(\sqrt{x^2+\rho_{\mathrm{ch}}^2/4T^2}-\omega_0/2T)}-\frac{\tanh\sqrt{x^2+\rho_{\mathrm{ch}}^2/4T^2}+\tanh(\omega_0/2T)}{2(\sqrt{x^2+\rho_{\mathrm{ch}}^2/4T^2}+\omega_0/2T)}\right). \tag{25}$$

(Here it was assumed that in the concentration region under consideration the ratio $\mu/T_\rho\simeq\epsilon_F/T_\rho\gg 1$ because $\mu\simeq\epsilon_F$.[10–12,16])

Since usually $\omega_0/T_\rho\gg 1$, only very small x give the main contribution to the integral (25) (this is seen from the limit $\rho/T_\rho\to 0$, when $\epsilon_F/\omega_0\to\infty$). Therefore, it takes the approximate form

$$\frac{2\pi}{g_{\rm ph}^2 m}=\int_0^\infty dx\left(\frac{\tanh\sqrt{x^2+\rho_{\rm ch}^2/4T^2}}{\sqrt{x^2+\rho_{\rm ch}^2/4T^2}}-\frac{1}{x+\omega_0/2T}\right). \quad (26)$$

On the other hand, the accepted condition $\rho_{\rm ch}(T_\rho)=0$ in (26) directly results in the simple equation

$$\frac{2\pi}{g_{\rm ph}^2 m}=\int_0^\infty dx\left(\frac{\tanh x}{x}-\frac{1}{x+\omega_0/2T_\rho}\right) \quad (27)$$

for T_ρ. Comparing (26) and (27), we obtain

$$\int_0^\infty dx\left(\frac{\tanh x}{x}-\frac{\tanh\sqrt{x^2+\rho_{\rm ch}^2/4T^2}}{\sqrt{x^2+\rho_{\rm ch}^2/4T^2}}\right)=\ln\frac{T_\rho}{T}.$$

Now from the expansions

$$\frac{\tanh\sqrt{x^2+\rho_{\rm ch}^2/4T^2}}{\sqrt{x^2+\rho_{\rm ch}^2/4T^2}}\simeq\begin{cases}1-3^{-1}[x^2+\rho_{\rm ch}^2/4T^2], & x\leqslant 1\\ x^{-1}-\rho_{\rm ch}^2/8T^2x^3, & x>1\end{cases},$$

we obtain the expression which we need

$$\rho_{\rm ch}(T)\simeq 2.62T_\rho\sqrt{T_\rho/T-1}. \quad (28)$$

Recall that the well-known 3D result is $\Delta_{\rm BCS}(T)=3.06T_{\rm BCS}^{\rm MF}\sqrt{T_{\rm BCS}^{\rm MF}/T-1}$ (Ref. 17) and the small difference can be explained by the above approximation, which is suitable for the qualitative discussion below (see Sec. 4).

The dependence (28) must be substituted in Eq. (22); because $\mu/T_{\rm BKT}\simeq\epsilon_F/T_{\rm BKT}\gg 1$ and $\rho_{\rm ch}(T_{\rm BKT})/T_{\rm BKT}\ll 1$ when $T_{\rm BKT}\to T_\rho$, this equation can be written as

$$\frac{\epsilon_F}{4T_{\rm BKT}}\left[1-\frac{\rho_{\rm ch}^2}{4T_{\rm BKT}^2}\frac{\partial}{\partial(\rho_{\rm ch}^2/4T_{\rm BKT}^2)}\right]\times\int_0^\infty dx\left(\frac{1}{\cosh^2 x}-\frac{1}{\cosh^2\sqrt{x^2+\rho_{\rm ch}^2/4T_{\rm BKT}^2}}\right)=1. \quad (29)$$

Finally, using the expansion in $\rho_{\rm ch}(T_{\rm BKT})/2T_{\rm BKT}$ in integral (29), the latter can be transformed to

$$\frac{a\epsilon_F}{8T_{\rm BKT}}\left[\frac{\rho_{\rm ch}(T_{\rm BKT})}{2T_{\rm BKT}}\right]^4=1, \quad (30)$$

where the numerical constant

$$a=\int_0^\infty dx\,\frac{\tanh^2 x-x^{-1}\tanh x+1}{2x^2\cosh x}\simeq 1.98.$$

Combining now (28) and (30), we obtain the final simple relation between T_ρ and $T_{\rm BKT}$ for the large carrier density

$$T_{\rm BKT}\simeq T_\rho(1-2.34\sqrt{T_\rho/\epsilon_F}).$$

In other words, $T_{\rm BKT}$ as a function of n_f (see Fig. 2) actually approaches T_ρ (or $T_{\rm BCS}^{\rm MF}$).

With regard to the crossover region defined by the equality $\mu\simeq 0$, it is easy to see from Eqs. (21)–(24) and Fig. 2 that the former corresponds to the densities when the temperatures T_ρ and $T_{\rm BKT}$ are essentially different; otherwise, the pseudogap phase really exists here. It is important that because of the relatively low, for the phonon case, value of the energy of state of the bound pair $\varepsilon_b^{\rm (ph)}$ and the very small region of negative μ (Ref. 16), the behavior $T_{\rm BKT}(n_f)\sim\epsilon_F$ hardly corresponds to the Bose–Einstein condensation and, in fact, is consistent for carrier densities when $\mu>0$ (although probably $\mu\neq\epsilon_{\rm F}$).

3) $g_{4\rm F}\neq 0$, $g_{\rm ph}\neq 0$.

This general case contains the boson exchange and unretarded interactions. The situation closest to the real systems corresponds to the case $g_{4\rm F}<0$ (or to some sort of short-range repulsion) but total interaction has attractive character; this means that at least $g_{\rm ph}^2>|g_{4\rm F}|$. There are two qualitatively different cases again: i) low and ii) high carrier densities.

i) $\epsilon_{\rm F}/\omega_0\ll 1$. For this inequality we see that T_ρ satisfies the same equation $T_\rho\ln(\epsilon_{\rm F}/T_\rho)=-\varepsilon_b/2$, where now

$$\varepsilon_b=-2W\left(\frac{\omega_0}{W+\omega_0}\right)^{\lambda/(\lambda-\mu_C^*)}\exp\left(-\frac{2}{\lambda-\mu_C^*}\right)$$

is the two-body bound state energy, and $\lambda=g_{\rm ph}^2 m/2\pi$ and $\mu_C^*=-g_{4\rm F}m/2\pi$ are the ordinary, effective, electron-phonon (attractive) and Coulomb (repulsive) constants; the difference $\lambda-\mu_C^*$ should be positive. We see that ε_b equals to $\varepsilon_b^{\rm (4F)}$ or $\varepsilon_b^{\rm (ph)}$ (if $W\gg\omega_0$) for the previous limiting cases. Here $T_{BKT}=\epsilon_{\rm F}/8$ and $\partial T_\rho(n_f)/\partial n_f\to\infty$ as $n_f\to 0$.

ii) $\epsilon_{\rm F}/\omega_0\gg 1$. In this limit the expression for T_ρ has the form

$$T_\rho=\frac{\gamma}{\pi}\sqrt{\omega_0|\varepsilon_b|\left(\frac{\epsilon_{\rm F}}{\omega_0}\right)^{\mu_C^*/(\lambda-\mu_C^*)}}. \quad (31)$$

It follows from this expression that the dependence on n_f is still weak, which results in an increase (for $g_{4\rm F}>0$) or decrease (for $g_{4\rm F}<0$) of T_ρ when $\epsilon_{\rm F}\gg\omega_0(\lambda-\mu_C^*)/\mu_C^*$, which is direct consequence of the model with unretarded 4F interaction. The temperature $T_{\rm BKT}$ is described by Eq. (30) with T_ρ, defined by (31).

It must be noted, however, that with more realistic assumptions about Coulomb repulsion μ_C^*, which can be initially represented (see, for example, Ref. 22) by the matrix elements

$$V(\mathbf{k},\mathbf{k}')=\begin{cases}V_c, & |\xi(\mathbf{k})|,|\xi(\mathbf{k}')|\leqslant|\mu|\\ 0, & |\xi(\mathbf{k})|,|\xi(\mathbf{k}')|>|\mu|\end{cases},$$

it acquires the well-known Tolmachev logarithmic correction or turns out to be screened at large n_f. At low carrier densities such effect (screening) does not take place, so the local repulsion model can be considered as a good approximation in the physical cases in which the Fermi energy of free fermions is less than or not much greater than the characteristic boson frequency.

4. CONCLUSIONS

The model proposed to describe the possible two-stage superconducting phase transition in 2D (and quasi-2D) metallic systems was greatly simplified in order to investigate their most typical and general features. Surprisingly, it gives some essential details which are characteristic of underdoped

HTSC copper oxides. In particular, the experimental data show[30,31] that i) the critical temperature T_c for low n_f indeed is proportional to n_f (which is simply ϵ_F), ii) T_c ''becomes saturated'' when n_f approaches ''optimal doping'' (i.e., carrier concentration when T_c as a function of n_f reaches its highest possible value for the given compound), iii) the ratio T_c/ϵ_F in these and other ''exotic'' superconductors is as high as $10^{-2}-10^{-1}$, which independently points to rather low Fermi energy, etc. (for details see Ref. 31). In addition, the standard ratio (see Ref. 23) $2\Delta(0)/T_c$ can be roughly estimated as $2\rho(0)/T_{BKT}$; this value always exceeds its canonical BCS value and increases approximately $\sim n_f^{-1/2}$ at small values of n_f.

One would think that the peculiarities mentioned above receive their natural interpretation on the basis of the model for the metal with indirect fermion-fermion interaction if the temperature T_{BKT} is the critical temperature T_c (this is justifiable for pure 2D systems[13]). In a quasi-2D model the third spatial direction and the phase fluctuation stabilization give rise to the true temperature T_c of an ordinary homogeneous ordering[31,32] (see also Ref. 11), but the region where $T_c \neq T_\rho$ (or T_c^{BCS}) can be conserved.[11]

As regards the other temperature (here estimated as T_ρ), it is usually determined empirically as some temperature point T^*, where the observable spectral (or magnetic) properties of HTSCs begin to deviate appreciably from their standard for normal metallic state behavior.[5–9] As a rule, such a deviation is attributable to the appearance of fluctuating (short-lived) pairs. We showed, however, that a finite number of these pairs does exist or begins to be formed (rapidly) below some definite (in the mean-field approximation) temperature T_ρ, which, as indicated above, does not correspond to a phase transition. Additionally, because of the fluctuations (including quantum fluctuations), $\rho(T_\rho)$ remains nonzero at $T>T_\rho$. In this temperature region the number of pairs is exponentially small, and the fluctuations, which are superconducting (developed in the 2D case) can contribute to the temperature behavior of different observables [even at large n_f (Ref. 33)]. The only difference from the supposed dependence T^* on the density of doped holes is the decreasing asymptotic behavior at n_f. We have found that this collective temperature also decreases, while (see, for example, Ref. 31) T^* is usually plotted as one that increases with decreasing n_f. It seems that such a behavior still has no satisfactory explanation, especially for the 2D case, where the bound states do not demand, as in the 3D case, a strong coupling. Nevertheless, it must be stressed that the above limit, $T_\rho(n_f)\to 0$ as $n_f\to 0$, cannot be considered as sufficiently regular because of the growth of the neutral OP fluctuations; their role is disregarded, and they become very important at small values of n_f, when, for example, any collective behavior cannot exist.

The model under consideration qualitatively correctly describes the explicit narrowing of the pseudogap area as the carrier density increases [such a narrowing results in a rather rapid confluence of the temperatures (T_c and T_ρ) and their experimental confluence, rendering them indistinguishable in the BCS limit]. On the other hand, recent angle-resolved photoemission spectra unexpectedly showed[2,3] that, in contrast with T_c, the superconducting gap even in the underdoped samples is essentially independent of doping. Such a difference to some extent also follows from the superconducting transition scenario proposed by us: Indeed, one-particle spectrum gap as a function of n_f is simply defined by the value $\rho(n_f)$ (it was calculated in Ref. 16), which is proportional to $T_\rho(n_f)$, and the latter (see Fig. 2) very quickly becomes equal to Δ_{BCS} or a constant, although T_{BKT} (and T_c) not quite yet reach this point. This behavior is a direct consequence of the evident smallness of the negative μ (local pairs and/or strongly developed fluctuations) region for the indirect fermion-fermion interaction model in which the bound states prove to be extremely subtle.

Some important problems still remain unresolved and must be investigated. These problems are: more complete and deep development of the model, which must consider different kinds of dispersion relations for the intermediate bosons; more careful allowance for the Coulomb repulsion; neutral OP fluctuations, especially for low n_f; generalization of the approach to the case of nonisotropic pairing. On the other hand, high-T_c compounds must be studied in the frame of more realistic models, which include such pecularities of HTSCs as the magnetism of cuprate layers, non-qudratic free carrier dispersion relation with possible van Hove singularities in the hole density of states, and, of course, spatial quasi-two-dimensionality. One of the most interesting problems is to obtain doping and temperature-dependent effective action, which is equivalent to the Ginzburg–Landau potential, because in many cases the phenomenology is more preferable.

We would like to thank V. P. Gusynin, S. G. Sharapov, and I. A. Shovkovy for extremely interesting and useful discussions concerning the questions raised in this study.

*)E-mail: vloktev@gluk.apc.org

1)It should be noted that Efetov and Larkin, in fact, were the first to use such a parametrization. They studied[27] the effect of interchain hopping and OP phase fluctuations on the superconducting transition temperature in 1D superconductors.

2)Equations for ρ_{ch} and ρ_{ins} are obtained below and, as was shown in Ref. 16, it is an admissible approximation to put in them the value ρ_{ch} (and ρ_{ins}), which is independent space and time variables.

3)Because ρ_{ch} and ρ_{ins} cannot exist simultaneously [see Eq. (23)], the index ρ is the only OP, which appears at a definite sign of the kernel (10).

4)Despite the fact that the temperature T_ρ is not identical to the BCS critical temperature T_{BCS}^{MF}, they coincide for the large carrier density only (see below).

5)Being equal (in the mean-field approximation only), these temperatures (T_ρ and T_{BCS}^{MF}) are in fact different: if T_{BCS}^{MF} immediately decreases to zero as the fluctuations ϕ and ϕ^0 are taken into account, T_ρ does not decrease and is renormalized only when ρ fluctuates.

[1] V. M. Loktev, Fiz. Nizk. Temp. **22**, 3 (1996) [Low Temp. Phys. **22**, 1 (1996)].
[2] D. S. Marshall, D. S. Dessau, A. G. Loeser *et al.*, Phys. Rev. Lett. **76**, 4841 (1996).
[3] H. Ding, T. Yokoya, J. C. Campuzano *et al.*, Nature (London) **382**, 51 (1996); H. Ding, M. R. Norman, T. Yokoya *et al.*, Phys. Rev. Lett. **78**, 2628 (1997).
[4] A. G. Loeser, Z.-X. Shen, D. S. Dessau *et al.*, Science **273**, 325 (1996).
[5] B. Gross-Levi, Prog. Theor. Phys. **49**, 17 (1996).
[6] B. Batlogg and V. J. Emery, Nature (London) **382**, 20 (1996).
[7] N. P. Ong, Science **273**, 321 (1996).

[8] C. Berthier, M. Julien, M. Horvatic, and Y. Berthier, J. Phys. I **6**, 2205 (1996).
[9] D. Pines, Tr. J. of Physics **20**, 535 (1996).
[10] M. Randeria, in *Bose–Einstein Condensation*, ed. by A. Griffin, D. W. Snoke, and S. Stringari (Cambridge University Press, N.Y. (1995), p. 355.
[11] V. M. Loktev and S. G. Sharapov, e-mail cond-mat/9706285 (1997); Cond. Mat. Phys. No. 11, 131 (1998).
[12] E. V. Gorbar, V. P. Gusynin, and V. M. Loktev, Supercond., Phys. Chem. Technol. **6**, 375 (1992); Fiz. Nizk. Temp. **19**, 1171 (1993) [Low Temp. Phys. **19**, 832 (1993)].
[13] R. MacKenzie, P. K. Panigrahi, and S. Sakhi, Int. J. Mod. Phys. A **9**, 3603 (1994).
[14] M. Drechsler and W. Zwerger, Ann. Phys. **1**, 15 (1992).
[15] V. P. Gusynin, V. M. Loktev, and S. G. Sharapov, JETP Lett. **65**, 182 (1997).
[16] V. M. Loktev, V. M. Turkowski, and S. G. Sharapov, J. Phys. Stud. **3**, 311 (1997).
[17] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics*, Prentice-Hall, Englewood, NJ (1963).
[18] E. A. Pashitskii, Sov. Low. Temp. Phys. **21**, 995 (1996); **21**, 1019 (1996).
[19] V. L. Ginzburg, Usp. Fiz. Nauk **167**, 229 (1997).
[20] Y. Nambu, Phys. Rev. **117**, 648 (1961).
[21] T. Kugo, Preprint-KUNS 1086 HE(TH)91/12 July, Kyoto, Japan (1991).
[22] S. V. Vonsovskii, Yu. A. Izyumov, and E. Z. Kurmaev, *Superconductivity of Transition Metals, Their Alloys and Compounds*, Nauka, Moscow (1977).
[23] J. R. Schrieffer, Theory of superconductivity, Benjamin, NY (1964).
[24] B. T. Geilikman, Usp. Fiz. Nauk **88**, 327 (1966) [Sov. Phys. Usp. **9**, 142 (1966)].
[25] J. M. Cornwall, R. Jackiw, and E. Tomboulis, Phys. Rev. D **10**, 2428 (1974).
[26] L. J. R. Aitchison, P. Ao, D. J. Thouless, and X.-M. Zhu, Preprint CERN-TH.7385/94 (1994).
[27] K. B. Efetov and A. I. Larkin, Zh. Eksp. Teor. Fiz. **66**, 1129 (1974) [Sov. Phys. JETP **39**, 1551 (1974)].
[28] A. A. Abrikosov, Phys. Rev. B **55**, R6149 (1997).
[29] Yu. A. Izyumov and Yu. N. Skryabin, *Statistical Mechanics of Magnetically Ordered Systems*, Nauka, Moscow (1987).
[30] Y. J. Uemura, Nature (London) **364**, 605 (1993).
[31] Y. J. Uemura, e-mail cond-mat/9706151 (1997).
[32] E. V. Gorbar, V. M. Loktev, and S. G. Sharapov, Physica C **257**, 355 (1996).
[33] V. M. Loktev and S. G. Sharapov, Fiz. Nizk. Temp. **23**, 180 (1997) [Low Temp. Phys. **23**, 132 (1997)].

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

Magnetoluminescence of $Ge/Ge_{1-x}Si_x$ heterostructures

A. V. Chernenko*)

Institute of Solid State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

N. G. Kalugin and O. A. Kusnetsov

Institute of Microstructure Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia
(Submitted 30 October 1997)
Zh. Éksp. Teor. Fiz. **114**, 619–627 (August 1998)

This paper reports on the first investigation made of luminescence of $Ge/Ge_{1-x}Si_x$ heterostructures at liquid-helium temperatures in a magnetic field of up to 14 T. The luminescence lines observed in the spectra are due to both free and impurity bound excitons in Ge layers. The diamagnetic shift of the quasi-two-dimensional exciton has been measured. From the experimental data the size of the exciton has been estimated to be 75–90 Å. © *1998 American Institute of Physics.* [S1063-7761(98)01808-3]

1. INTRODUCTION

The interest in studies of optical properties of two-dimensional Si/Ge structures is stimulated primarily by the prospect of integrating them in traditional silicon microdevices and using them as sources and detectors of light.[1] Moreover, high-quality structures based on Si/Ge are of great interest in view of studying many-body effects in the system of photogenerated excitons. Since radiative recombination is slower than in structures based on III–V semiconductors, one can expect longer exciton lifetimes ($\tau \sim 10^{-5}-10^{-6}$ s). An electron–hole system with longer τ should be capable of cooling down to a temperature close to that of the crystal lattice. In this connection, we note a recent publication[2] reporting an observation of a line corresponding to recombination of quasi-two-dimensional excitonic molecules (biexcitons) in luminescence spectra of a Si/SiGe/Si quantum well.

Since the publication by Sturm *et al.*,[3] who were the first to observe excitonic luminescence from $Si/Si_{1-x}Ge_x/Si$ quantum wells, the number of investigations of luminescence in such structures has been quite considerable. The nature and properties of lines in these spectra have been studied in detail. At the same time, only two publications[4,5] were dedicated to luminescence of Ge/GeSi multiple quantum well (MQW) structures. Luminescence lines detected in spectra of Ge/GeSi MQW structures were attributed to annihilation of free and, probably, impurity bound excitons in Ge layers, assuming that these features could not be resolved. On the low-energy side of the dominant spectral line, a marked ''shoulder'' (the $x1$ line in Fig. 1; see also Ref. 5) was detected, whose nature has remained unclear and which was tentatively ascribed to an excitonic state with an energy lower than that of the free exciton, such as an excitonic molecule.

Earlier studies[6] of excitonic molecules in uniaxially strained Ge indicated that, under the condition $\tau_{\rm exc} \gg \tau_s$ (where $\tau_{\rm exc}$ is the exciton lifetime and τ_s is its spin relaxation time), an excitonic molecule is destroyed by a magnetic field higher than $B_c = \Delta_M / g\mu_B$, where Δ_M is the biexciton binding energy. In uniaxially strained Ge the excitonic g-factor is considerable, and dissociation of a molecule induced by magnetic field occurs at $B_c \sim 1.5$ T. A similar biexciton behavior should have been expected in the case under investigation, which would allow us to reliably identify these species. Note that in strained silicon we have $\tau_{\rm exc} < \tau_s$, and excitonic molecules are not destabilized by external magnetic fields. It seems that this property also applies to the case of $Si/Si_{1-x}Ge_x/Si$ quantum wells. Since the diamagnetic shift is a function of the exciton size,[7] the shift of excitons localized at impurities and on fluctuations of the random potential plotted against magnetic field should be smaller than that of free excitons. At an appropriate magnetic field strength, this should allow us to resolve different excitonic lines. Thus, magnetoluminescence measurements show promise in view of identifying different lines in luminescence spectra.

This paper reports on the first measurements of magnetoluminescence ever made in $Ge/Ge_{1-x}Si_x$ MQW structures.

2. EXPERIMENTAL PROCEDURE AND DISCUSSION

In our experiments we used $Ge/Ge_{1-x}Si_x$ structures grown by the vapor-phase hydride technique on Ge[111] substrates with the following parameters. Sample No. 1 had Ge layers with thickness $d_{\rm Ge} = 210$ Å, superlattice period $D = 530$ Å, number of periods in the sample $N = 72$, $x = 12.3\%$, and lattice constant in the lateral direction $a = 5.638$ Å, whereas $d_{\rm Ge} = 120$ Å, $D = 420$ Å, $N = 262$, $x = 14\%$, and $a = 5.629$ Å in sample No. 2. The residual concentration of n-impurities in Ge layers was $(1-2) \times 10^{13}$ cm^{-3}. A detailed description of the samples and manufacturing technique can be found in Ref. 5 and references therein, where samples No. 1 and No. 2 are labelled as No. 261 and No. 262, respectively. The pump radiation was generated by Nd:YAG and Ar^+ cw lasers operated at wavelengths of 1064 and 488 nm, respectively. Spectra were dis-

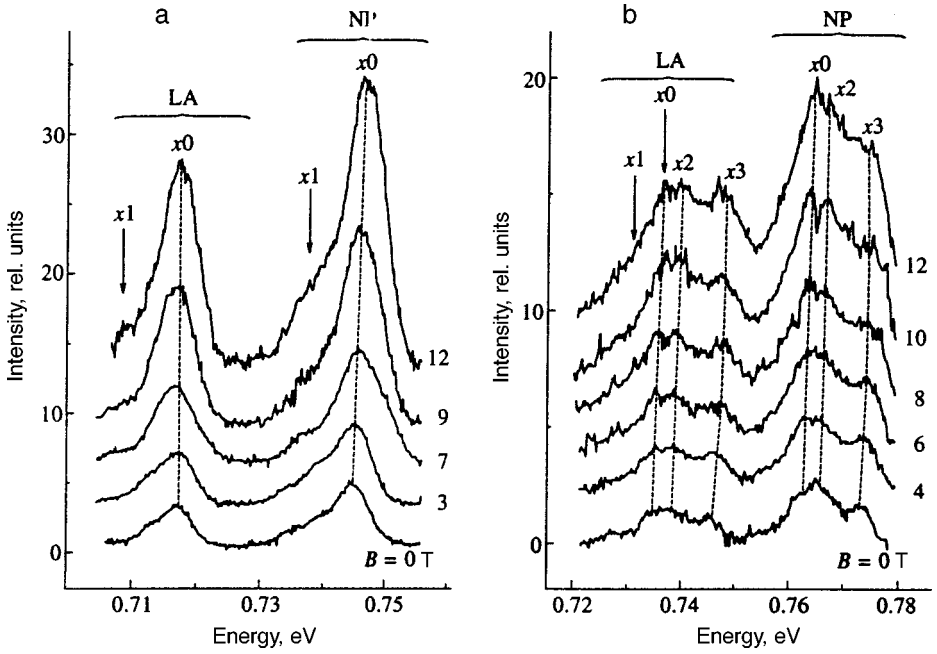


FIG. 1. (a) Nonphonon (NP) and LA-phonon components of luminescence spectra of sample No. 1 ($d_{Ge}=120$ Å) at liquid-helium bath temperature $T_b=2$ K, under cw pumping by the Nd:YAG laser at a power density of $W=30\,W/cm^2$ and at different magnetic fields. The shoulder on the low-energy side of the $x0$ line is labeled $x1$. (b) Luminescence spectra of sample No. 2 ($d_{Ge}=210$ Å for a constant Nd:YAG pump at a power density $W=55\,W/cm^2$, $T_b=2$ K, and at different magnetic fields. The features of the luminescence spectrum are labeled by $x0$, $x1$, $x2$, and $x3$, and their peaks are connected by the dashed lines.

persed by a grating monochromator and recorded using a cooled germanium light detector operated in the phase-locked mode. Measurements in a magnetic field were performed in a cryostat with a superconducting solenoid in the Faraday configuration. The pump radiation was fed to the sample via an optic fiber, and luminescence from the sample was picked up by the same fiber. In measurements at zero magnetic field, an optical cryostat was used.

Energy spectra of electrons and holes in $Ge/Ge_{1-x}Si_x$ heterostructures were calculated by Aleshkin and Bekin.[8] These calculations indicate that Ge layers are quantum wells for both electrons and holes. Positions of lines in luminescence spectra are in reasonable agreement with calculations.[5] The lattice constants of materials in the quantum wells and barriers (Ge and GeSi) are notably different (up to $\sim 0.4\%$), which generates strain in the two-dimensional layers. Ge layers are compressed in the lateral directions, which can be represented as a hydrostatic compression combined with uniaxial tensile strain along the [111] axis. Using the formulas from Ref. 8, we find that Ge layers in sample No. 2 ($d_{Ge}=120$ Å) are under an equivalent uniaxial stress $P\simeq 8.5$ kbar, and in sample No. 1 ($d_{Ge}=210$ Å) $P\simeq 5.8$ kbar. In the presence of uniaxial stress ($\Delta = dS_{44}/\sqrt{3}\simeq 3.6$ meV/kbar, where d is the deformation potential and S_{44} is the compliance coefficient[9]) and size quantization, the splitting between the subbands in the valence band is 20–30 meV.[8]

Figure 1 shows luminescence spectra of heterostructures No. 1 and 2 (labelled by a and b, respectively) at high pump power and in different magnetic fields aligned normally to the superconducting layers. The intensity of luminescence lines increases with the magnetic field B up to a factor of about two at $B=12$ T. In both structures, the high-energy sideband of the luminescence line slightly shifts toward high energies.

It is clear that the intensity of the sideband on the low-energy side of the dominant line in sample No. 1 (it is denoted by $x1$ in Fig. 1a) rises with the magnetic field faster than that of line $x0$. The spectra in Fig. 1b, whose shapes are more complicated than the curves in Fig. 1a, show that features $x2$ and $x3$ transform to well resolved lines, and the intensities of the low-energy lines increase faster than those of lines corresponding to higher energies. At higher magnetic fields, a feature similar to $x1$ in Fig. 1a can be seen (it is also labelled as $x1$ in the other figures). In addition, at a fixed magnetic field, the intensities of different lines were found to depend differently on the pump power. One can see in Fig. 2 that the lower the line energy, the faster its growth with the pump power. At zero magnetic field, the intensities and po-

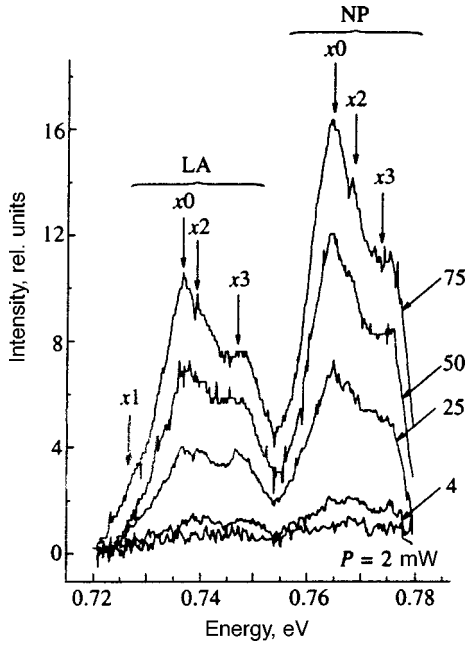


FIG. 2. Luminescence spectrum of sample No. 2 at $T_b=4.2$ K and a fixed magnetic field $B=12$ T at different powers of the Nd:YAG laser radiation.

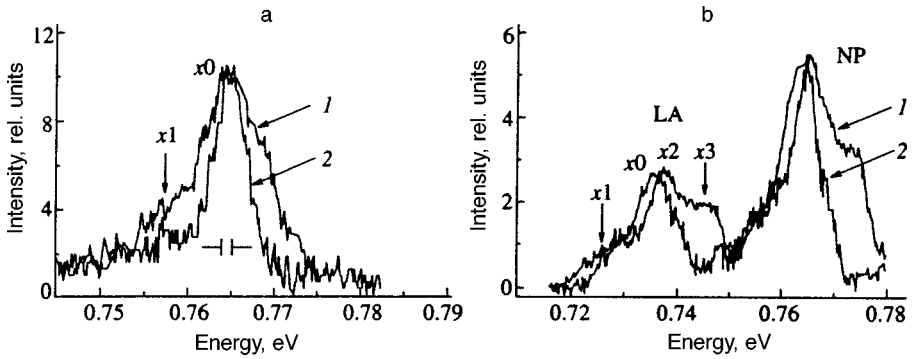


FIG. 3. (a) NP-components of luminescence spectra of sample No. 1 at $T_b=4.2$ K (*1*) excited by the Nd:YAG laser at $W=80$ W/cm^2 and (*2*) by the Ar^+ laser at $W=35$ W/cm^2. The spectra are normalized to the amplitudes of $x0$ line. (b) Luminescence spectra of sample No. 2 (*1*) before and (*2*) after etching. The spectra were taken at $T_b=4.2$ K under excitation by the Nd:YAG at a power density $W=50$ W/cm^2.

sitions of different lines change similarly with the temperature and pump power. The separation between the peaks of lines $x0$ and $x1$ in Fig. 1b is about 3.4 meV, and between the peaks $x0$ and $x2$ about 7.5 meV. When the laser spot was scanned over the sample surface, the positions and shapes of different lines remained constant, whereas changes in their relative intensities were within 15%.

When a sample was excited by radiation from the Nd:YAG laser (the penetration depth at $\lambda=1.064$ μm is about 1 μm and at $\lambda=488$ Å about 8 nm), the relative intensity of the $x1$ line with respect to $x0$ is higher than under Ar^+ laser radiation, as can be seen in Fig. 3a. This led us to conjecture that different lines in the spectra of both structures were due to luminescence from different quantum wells.

In order to test this conjecture, we etched sample No. 2 in the SR-4A solution for about one second. Figure 3b gives the spectra of sample No. 2 (curve *1*) before and (curve *2*) after the etching. One can see that the line $x2$ vanished from the spectrum after etching, and the width of the line $x0$ diminished. This result confirms our assumption.

The parameter characterizing different quantum wells in heterostructures is their width. The shift of the exciton line due to small changes in the quantum well width L_z is

$$\Delta E_{\rm exc}=|\delta E_{\rm exc}/\delta L_z|\Delta L_z=\pi^2\hbar^2\Delta L_z/(\mu_\parallel L_z^3),$$

$$1/\mu_\parallel=1/\mu_\parallel^e+1/\mu_\parallel^h,$$

where $\mu_\parallel^e$ and $\mu_\parallel^h$ are the reduced electron and hole masses in the vertical direction and m_0 is the free electron mass.[9] Hence the $x1$ and $x2$ lines correspond to luminescence from quantum wells near the surface whose widths differ by $\Delta L_z=4-6$ Å (2–3 monolayers) and 6–10 Å (3–4 monolayers), respectively.

The spectra of sample No. 1 (Fig. 1a) and of the etched sample No. 2 (Fig. 3b) are the simplest, and are therefore more convenient for measurements of the diamagnetic shift. Figure 4b shows luminescence spectra of the etched sample No. 2 in magnetic field aligned with the $[\bar{2}11]$ axis parallel to the layers (i.e., perpendicular to the [111] axis). Unlike the spectra of the same sample recorded in perpendicular magnetic field (Fig. 4a), they show the exciton luminescence intensity slowly changing with magnetic field B, but the shift of the high-energy sideband of the exciton line is comparable to that in the case $B\|[111]$. The spectra in the field aligned with the [011] axis are almost identical to those in Fig. 4b.

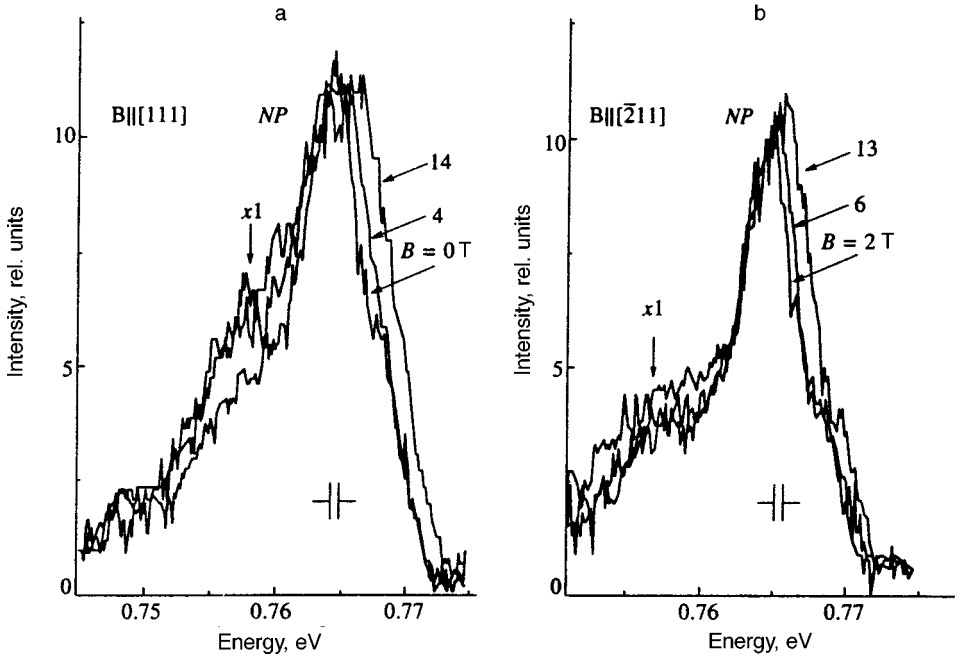


FIG. 4. Nonphonon (NP) components of sample No. 2 luminescence spectra after etching with cw Nd:YAG pumping at $T_b=4.2$ K in a magnetic field: (a) $B\|[111]$, pump power density $W=50$ W/cm^2, the spectra are normalized to the peak intensities of NP lines; (b) $B\|[\bar{2}11]$, pump pwer density $W=20$ W/cm^2.

The spectra of sample No. 1 are similar to the spectra of sample No. 2 in Fig. 4.

The exciton luminescence line in a perpendicular magnetic field is apparently enhanced because of the decrease in the diffusion length of photogenerated electron–hole pairs in a magnetic field, which should increase the probability of their localization on fluctuations of the random potential and irregularities of heterojunctions, thereby raising the radiative recombination probability.[10]

An estimate of the excitonic line diamagnetic shift using the point at half maximum on the high-energy edge yields $\sim$2 meV in a field of 14 T. The position of the similar point on the low-energy edge is, in fact, constant. The luminescence line broadens with the magnetic field because it is evidently a complex line incorporating lines of both free and bound excitons. Then the shift of the high-energy edge in magnetic field is controlled by diamagnetic properties of free quasi-two-dimensional excitons.

In order to test this assumption, let us estimate the shift of the excitonic line in the magnetic field B.

The hole masses in both size quantization subbands are highly anisotropic. The lowest subband of heavy holes (hh) with the larger density-of-states mass has the following masses: along the [111] axis $m_{\parallel}^{hh}=0.49m_0$, and in the plane normal to the [111] axis $m_{\perp}^{hh}=0.053m_0$. In the subband of light holes (lh) we have $m_{\perp}^{lh}=0.13m_0$ and $m_{\parallel}^{lh}=0.049m_0$, where m_0 is the free electron mass.[9] The holes from the split subbands are characterized by definite projections j_z of the full angular momentum on the [111] axis. The holes of the lower subband have $j_z=\pm 3/2$. Under uniaxial strain, levels in the conduction band also split. The minimum of the conduction band is formed by three equivalent valleys, whereas the energy of the forth valley, which is symmetrical about the [111] axis, shifts toward higher energies at a rate $\simeq$11 meV/kbar. Notwithstanding the high anisotropy of both electron and hole masses, the excitonic mass anisotropy in the $Ge/Ge_{1-x}Si_x$ quantum well is considerably lower. The energy surfaces for electrons and holes are ellipsoids of revolution, but their symmetry axes are different ([111] for holes, and [11 1] and equivalent axes for electrons). The longer axis of the reduced mass ellipsoid for an exciton incorporating an electron from any of the equivalent valleys makes an angle of about 21° with the [111] axis. The exciton's reduced masses are $\mu_1=0.045m_0$, $\mu_2=0.032m_0$, and $\mu_3=0.087m_0$. The excitonic masses in Ge stretched along the [111] axis are very close to those in Ge compressed along the [100] axis: $\mu_1'=0.047m_0$, $\mu_2'=0.031m_0$, and $\mu_3'=0.083m_0$. This allows us to use variational calculations of the exciton binding energy Ry in strained Ge[100][11] and obtain fairly accurate results. These calculations yield Ry$\simeq$2.8 meV. The exciton sizes corresponding to the reduced masses given above are $a_1\approx 170$ Å, $a_2\approx 200$ Å, and $a_3\approx 140$ Å.

A simple analytical calculation which takes into account barrier penetration of electron and hole wave functions and is in reasonable agreement with exact calculations of E_0 for a GaAs/AlGaAs quantum well[12] yields $E_0\approx 2.1$ Ry$\simeq$5.9 meV for a well with $d_{Ge}=120$ Å and $E_0=1.8$ Ry$\simeq$5 meV for a well with $d_{Ge}=210$ Å.

The exciton's diamagnetic susceptibility in bulk Ge is a function of electron and hole masses and, unlike the binding energy, is different for Ge compressed along the [100] axis and stretched along the [111] axis. But since the reduced masses in these two cases are quite close, calculations[11] of χ_{ij} for compressed Ge[100] and stretched Ge[111] are also close. This theory yields diagonal elements of the diamagnetic susceptibility tensor $\chi_{ii}\approx 0.3-0.5$ meV/T^2. Moreover, the Van Vleck correction to the diamagnetic susceptibility is only several percent of the Langevin susceptibility.[11]

The low-field diamagnetic shift of an isotropic quasi-two-dimensional exciton (either free, localized, or bound to an impurity) is described by an expression for the Langevin correction similar to that in the three-dimensional configuration[13]: $\Delta E=e^2B^2\langle r^2\rangle/(8\mu_\perp c^2)$, where $\langle r^2\rangle \equiv\langle\Psi|r^2|\Psi\rangle$ characterizes the exciton localization and $\mu_\perp$ is the reduced mass in the plane perpendicular to the magnetic field. In the isotropic case, the diamagnetic correction to the quasi-two-dimensional exciton energy is a factor of $\langle r^2\rangle/2a_B$ smaller than in the three-dimensional configuration. For an ideally two-dimensional exciton this ratio is 16/3. The exciton localization on irregularities of the potential and impurities also leads to a smaller ΔE. A value of $\chi_{\parallel}^{2D}$ several times smaller than in the three-dimensional case should be expected.

In order to calculate changes in the exciton energy due to magnetic field, one should also take into account the term linear in B: $\Delta E=(1/2)g_{exc}\mu_B B$ (g_{exc} is the exciton g-factor equal to the sum of electron and hole g-factors). Measurements of Shubnikov oscillations as functions of the magnetic field tilt angle in $Ge/Ge_{1-x}Si_x$ quantum wells selectively doped with boron and whose parameters are close to those of quantum wells studied in our experiments ($x=14\%$, $d_{Ge}=110$, 180 Å) indicate that the spin projection on the [111] axis in magnetic field is constant.[14] The g-factor derived from these measurements is -5.8. In the case of bulk Ge under a strong tensile strain along the [111] axis one has $g_h\simeq -7$.[9] For electrons of all three equivalent valleys in a magnetic field aligned with the [111] axis, $g_e\simeq 1$. In a magnetic field $B\|[211]$, g_e is different for different valleys. The diamagnetic shift was a maximum for excitons containing electrons from the [11 1] valley. These electrons have $g_e\simeq 1.5$.

Thus, the Zeeman component in the exciton energy for $B\|[\underline{1}11]$ is $\Delta E\simeq -(7/2)\mu_B B=-0.2B$ meV, and for $B\|[211]$ $\Delta E=-(1/2)\mu_B B\simeq -0.03B$ meV. The red shift in low magnetic fields, when the linear term dominates, is within 0.4 meV and could barely be detected in our experiments.

Since the diamagnetic shift in magnetic fields parallel and perpendicular to the structure layers is about 2 meV at $B=14$ T, by solving a set of two equations we obtain $\chi_\parallel=0.04$ meV/T^2 and $\chi_\perp=0.03$ meV/T^2. Thus, it turns out that the susceptibility χ of the two-dimensional exciton is a factor of 8–10 lower than that of the three-dimensional exciton. Hence, using the relationship[13] $\chi^{3D}/\chi_\parallel^{2D}=\langle r^2\rangle/2a_B^2$, we obtain the size of the region in which a quasi-two-dimensional exciton is localized: $\sqrt{\langle r^2\rangle}=0.4-0.5a_B=75-90$ Å.

The width of the excitonic luminescence line in quantum wells is determined by fluctuations in their widths and com-

position of the solid solution. In the case of excitons localized in wide Ge wells, when $L_z \geqslant a_B$, fluctuations in the solid solution composition can be neglected. If fluctuations in L_z are Gaussian, the luminescence line width adequately characterizes the random potential.[15]

In this case,[16] the exciton luminescence line has a Gaussian bell shape with FWHM

$$\sigma = 1.18\sqrt{p(1-p)}\,\delta_z \delta_{xy} \left| \delta E_{\text{exc}} / \delta L_z \right|_{L_z = L_0} / (2a_\perp),$$

where L_z is the well width and δ_z is the amplitude of fluctuations in L_z. Assuming that the amplitude of width fluctuations is within the spread of the widths of different wells, which was determined above, we take $\delta_z = 10$ Å. Let us take p, the average fraction of ''islands'' and ''depressions'' in the total interface area, to be equal to the most probable value of 0.5. The lateral exciton size in our estimate $a_\perp = \sqrt{\langle r^2 \rangle} = 80$ Å. The exciton luminescence line has the maximum width $\sigma_{\max}$ when the typical size of interface irregularities equals the lateral exciton size.[16] Then $\sigma = 0.59\delta_z \left| \delta E_{\text{exc}} / \delta L_z \right|_{L_z = L_0}$. In the case under consideration, $\sigma \simeq 2.5$ meV. When a sample is pumped by Ar^+ laser radiation, luminescence is generated only in several surface wells. The minimum width of measured luminescence lines proved to be 4.5–5 meV. Since the exciton binding energy on impurities in Ge layers is 1.5–2 meV,[5] the estimate of the exciton line width is in fair agreement with the measured σ.

Recent studies of exciton localization in strained InGaAs/GaAs structures indicated that the Stokes shift and luminescence line width are virtually independent,[17] which means that fluctuations in the random potential in such structures are non-Gaussian and caused by relaxation of elastic energy on heterojunctions.

It seems that this conclusion also applies to the strained $Ge/Ge_{1-x}Si_x$ structures studied in this work. In this case, although the estimate of σ given above is in agreement with measurements, the luminescence line width cannot be used as a unique characteristic of the random potential.

3. CONCLUSION

We have demonstrated that our observations can be interpreted in terms of diamagnetic properties of quasi-two-dimensional excitons which can be either free, or bound at impurities, or localized in Ge layers. Since the diamagnetic shift of a bound exciton is smaller than that of a free exciton, the broadening of luminescence lines in a magnetic field favors the hypothesis about its complex nature. Measurements of the exciton diamagnetic shift yield an estimate of the lateral exciton size. At the same time, we have obtained no indications of the existence of biexcitons in the studied structures.

We are indebted to Z. F. Krasil'nik and V. D. Kulakovskii for their involvement in this work and for illuminating discussions, and to V. B. Timofeev for useful remarks on the text of this paper.

The work was supported by the Russian Fund for Fundamental Research (Projects 29-02-17535 and 96-02-16991) and *Physics of Nanostructures* program (Projects 97-1065 and 96-2011).

*)E-mail: chernen@issp.ac.ru

[1] U. König, Phys. Scr. **68**, 90 (1996).
[2] T. W. Steiner, L. C. Lenchyshyn, M. L. W. Thewalt *et al.*, Solid State Commun. **89**, 429 (1994);M. L. W. Thewalt, *Extended Abstracts of 1993 Int. Conf. on Solid State Devices and Materials* (SSDM-93), Makuhari, Japan (1993), p. 494.
[3] J. C. Sturm, H. Monoharan, L. C. Lenchyshyn *et al.* Phys. Rev. Lett. **66**, 1362 (1991).
[4] N. G. Kalugin, L. K. Orlov, and O. A. Kuznetsov, JETP Lett. **58**, 200 (1993).
[5] L. K. Orlov, V. Ya. Aleshkin, N. G. Kalugin *et al.*, J. Appl. Phys. **80**, 415 (1996).
[6] V. D. Kulakovskii, V. G. Lysenko, and V. B. Timofeev, Usp. Fiz. Nauk **147**, 3 (1985) [Sov. Phys. Usp. **28**, 735 (1985)].
[7] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Pergamon Press, Oxford (1977).
[8] V. Ya. Aleshkin and N. A. Bekin, Fiz. Tekh. Poluprovodn. **31**, 171 (1997) [Semiconductors **31**, 132 (1997)].
[9] G. E. Bir and G. L. Pikus, *Symmetry and Deformation Effects in Semiconductors*, Wiley, New York (1974); Russian orig. Nauka, Moscow (1972).
[10] K. Fujii, T. Tomaru, T. Ohayama, and E. Otsuka, in *High Magnetic Field in Semiconductor Physics*, ed. by G. Landwehr, Springer-Verlag (1989), p. 558.
[11] T. G. Tratas and V. M. Édel'shtein, Zh. Éksp. Teor. Fiz. **81**, 696 (1981) [Sov. Phys. JETP **54**, 372 (1981)].
[12] H. Mathieu, P. Lefebre, and P. Christol, Phys. Rev. B **46**, 4092 (1992).
[13] W. Ossau, B. Jäkel, E. Bangert *et al.* Surf. Sci. **174**, 188 (1986).
[14] N. A. Gorodilov, O. A. Kuznetsov, L. K. Orlov *et al.*, JETP Lett. **56**, 394 (1992).
[15] F. Yang, M. Wilkinson, E. J. Austin *et al.*, Phys. Rev. Lett. **70**, 323 (1993).
[16] J. Singh and K. K. Bajaj, J. Appl. Phys. **57**, 5433 (1985).
[17] W. Broun, L. V. Kulik, T. Baars *et al.*, submitted to Phys. Rev. B (1998).

Translation provided by the Russian Editorial office.

Amplification of spin oscillation noise by self-modulation in a traveling magnetostatic wave

L. L. Savchenko, S. A. Nikitov*), A. F. Popov, and M. V. Chetkin

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 103907 Moscow, Russia
(Submitted 13 November 1997)
Zh. Éksp. Teor. Fiz. **114**, 628–639 (August 1998)

A theoretical analysis is made of the formation and evolution of modulational instability of a traveling magnetostatic spin wave in a ferromagnetic film. Results of earlier experiments are analyzed and it is demonstrated that the experimental results can be explained using a model describing the evolution of the signal wave and initial noise allowing for their nonlinear interaction. The instability process is simulated numerically and the results are compared with calculations using a deterministic model. Mechanisms are discussed for the loss of spectral symmetry in the formation of modulation frequency satellites. © *1998 American Institute of Physics.* [S1063-7761(98)01908-8]

1. INTRODUCTION

Yttrium iron garnet films are important for studying the physical properties of traveling magnetostatic spin waves, which can easily be excited and propagate readily in these films. As the power of the oscillator exciting the signal wave increases, the wave may become unstable against modulation of the wave amplitude. The instability of magnetostatic spin waves in magnetic films has been studied in Refs. 1–11. In Refs.1–4, for example, first-order decay processes in yttrium iron garnet films were observed experimentally in the phase-matching region for three interacting waves (three-magnon processes). Outside these regions four-wave scattering processes predominate, for which the phase-matching conditions have the form

$$2\omega(\mathbf{k}_0)=\omega_1(\mathbf{k}_1)+\omega_2(\mathbf{k}_2), \quad 2\mathbf{k}_0=\mathbf{k}_1+\mathbf{k}_1. \tag{1}$$

This type of process for magnetostatic spin waves traveling in magnetic films was observed in Refs. 5–11. This effect is characterized by the appearance of satellite signals in the spectrum of the transmitted wave, equidistant in frequency and wave number from the initial wave. As the electromagnetic pump power increases, one of the satellites is replaced by a high-intensity noise signal with a broad region of excitation spectrum.[5–7] This noise signal was interpreted in Refs. 5 and 6 as a kinetic instability caused by the secondary generation of spin waves.[12]

In addition to the phase-matching conditions (1) being satisfied, the onset of parametric instability requires the wave power to exceed some threshold determined by the losses accompanying the wave propagation. The parametric instability may be satisfied for several pairs of waves simultaneously. If the group velocities of the interacting waves have similar values and directions, parametric instability is converted into a self-modulational instability in which competition between the nonlinear and dispersion shifts of the wave phase velocities plays a major role. Self-modulation effects involving the self-interaction of magnetostatic spin waves in magnetic films caused by the tendency to soliton formation were discussed in Refs. 13–17. According to theory, the second-order parametric scattering thresholds are determined by the condition[18]

$$\nu=\sqrt{\beta_{12}^2\varphi^4-\frac{1}{4}[\omega_1(\mathbf{k}_1)+\omega_2(\mathbf{k}_2)-2\omega(\mathbf{k}_0)]^2}>\delta\omega, \tag{2}$$

where ν is the instability growth rate, $\delta\omega$ is the ferromagnetic resonance line width for a particular wave, which determines the losses in the wave, β_{12} is the coefficient of parametric coupling of the harmonic formed as a result of decay of the fundamental-frequency wave, and φ is the amplitude of the signal wave (angular deflection). As this wave decays into waves with a large difference between the group velocities v_{gi}, when the frequency detuning is determined mainly by the linear sections of the dispersion curve, i.e.,

$$\Delta\omega=\frac{1}{2}(v_{g1}+v_{g2}-2v_{g0})\Delta k,$$

the dispersion phase shift does not play a significant role. The growth rate has a maximum at the phase-matching point $\Delta\omega=0$ since

$$\nu=\sqrt{\beta_{12}^2\varphi^4-(\Delta\omega)^2}. \tag{3}$$

In the modulational instability the growth rate depends on the dispersion and nonlinear frequency shifts of the interacting waves, since the linear detuning as a result of a difference between the group velocities plays a minor role in this mechanism. The growth rate of this type of instability is given by

$$\nu=\sqrt{-\frac{\omega_{kk}}{v_g^2}(\Delta\omega)^2\left[\frac{\omega_{kk}}{v_g^2}(\Delta\omega)^2+2\beta_{00}\varphi^2\right]}, \tag{4}$$

where ω_{kk} is the dispersion coefficient and β_{00} is the coefficient of nonlinear frequency shift. From this it follows that the maximum growth rate of the self-modulational perturbation is achieved when

$$\Delta\omega_0=\sqrt{|2\beta_{00}(v_g^2/\omega_{kk})\varphi^2|}$$

and is proportional to the signal wave amplitude. For this last case, the spectral width of the region in which the decay perturbations occur, determined from their growth rate, differs substantially from the previous case. Whereas in the first case this width is determined only by the nonlinearity $\Delta\omega_{\max}=\beta_{12}\varphi^2$, in the second case it has a root dependence on the dispersion and the nonlinear frequency shift, since $\Delta\omega_{\max}=2\Delta\omega_0$. In view of this, the peak width of the self-modulational satellite may exceed by an order of magnitude or more the frequency width of the peak of the parametrically excited wave in the magnetic film having a different group velocity. In analyses of wave decay processes in films this factor has not usually received any attention and the appearance of broad frequency peaks has been attributed to the onset of a ''kinetic'' instability. If we bear in mind that as the pump wave propagates it undergoes damping as a result of dissipation, the relative gain of the wave satellite will also depend on the propagation length. The spectral dependence of the self-modulational peak in the evolution of the interacting waves in a dissipative nonlinear medium was analyzed using a model of the generalized nonlinear Schrödinger equations for light propagation in a fiber.[19] Numerical calculations of the evolution of instability in a magnetostatic wave caused by self-interaction effects were reported in Refs. 20 and 21. However, these studies contained no detailed spectral analysis of the modulated magnetostatic wave. In addition, it should also be noted that the evolution of the parametric instability in the nonlinear Schrödinger equation model is usually simulated using a harmonic initial perturbation of the amplitude of the exact solution at the initial point[19,22]

$$A(0)=A_0(1+\varepsilon\cos\omega t),$$

where $A(0)$ is the signal amplitude. In reality, however, a random perturbation exists at the input, corresponding to the noise spectrum of the exciting signal which is in fact the noise of the oscillator used to excite the rf current which in turn excites the magnetostatic waves.

Here we propose a model of self-modulational instability as a process involving the evolution of the signal wave and the noise, with allowance for their nonlinear interaction. We shall compare the results of this simulation of the instability of a traveling spin wave with the calculations using a deterministic model. In addition, we shall also discuss possible mechanisms for the loss of spectral symmetry in the formation of the satellites in modulational instability which are observed in experiments using magnetic films. Thus, our aim is to perform a theoretical analysis and numerical simulation of the formation and evolution of the self-modulation noise of spin oscillations in a traveling spin wave and to discuss mechanisms for the occurrence of spectrally broad noise signals and loss of spectral symmetry of the satellites formed in modulational instability, which are observed in experiments using magnetic films.

2. EQUATION FOR THE EVOLUTION OF THE ENVELOPE AND LINEAR ANALYSIS OF SELF-MODULATIONAL INSTABILITY

If the nonlinear dispersion dependence is known for the wave number $k=k(\omega,|\varphi|^2)$ where φ is the wave amplitude, the evolution of the wave packet in the Fourier representation

$$\varphi(z,t)=\exp[i(k_0z-\omega_0t)]\int\varphi_\Omega(z)\exp(-i\Omega t)d\Omega,$$

where $\Omega=\omega-\omega_0$ is the frequency detuning, may be written using the equation for the envelope[23]

$$\frac{\partial\varphi_\Omega}{\partial z}=[-\Gamma(\Omega)+ik(\Omega)]\varphi_\Omega + ik_{|\varphi|^2}[|\varphi(z,t)|^2\varphi(z,t)]_\Omega, \tag{5}$$

where $k(\Omega)=k(\omega,0)-k_0$ is the spectrum of a low-amplitude wave near the frequency and wave number of the signal wave, $k_{|\varphi|^2}=(\partial k/\partial|\varphi|^2)_{\varphi=0}$ is the coefficient of the nonlinear frequency shift, and $\Gamma(\Omega)$ is the damping factor. In the literature analyses are usually confined to second-order dispersion and the linear frequency shift is taken into account by introducing the space–time coordinate $z-v_gt$, traveling with the group velocity v_g. In this case, we have $k(-\Omega)=k(\Omega)$. We shall subsequently consider the more general case, $k(-\Omega)\neq k(\Omega)$, which specifically allows for third-order dispersion.

The dispersion parameters for various types of spin waves in a magnetic film of thickness d and the nonlinear coefficients are given in Refs. 13 and 24. For example, for a normally magnetized film and for the initial region of the excitation spectrum of the dominant mode of forward magnetostatic waves these parameters are

$$\Gamma(\Omega)=\frac{\delta\omega}{v_g},\quad v_g=\left(\frac{\partial k}{\partial\omega}\right)^{-1},$$

$$k(\Omega)=\frac{1}{d\sqrt{-\mu}}\arctan\left(-\frac{2\sqrt{-\mu}}{\mu+1}\right),$$

where

$$\mu=\frac{w_H^2+w_Hw_M-\Omega^2}{w_H^2-\Omega^2},$$

$$w_H=\gamma(H-4\pi M),\quad w_M=4\pi M\gamma,$$

γ is the magnetomechanical ratio, $4\pi M$ is the saturation magnetization, H is the external magnetic field, and $k_{|\varphi|^2}=w_M/v_g$.

Equation (5) has a solution in the form of a damped nonlinear wave at the carrier frequency ω_0 with $\varphi_\Omega=\Phi_0(z)\delta(\Omega)$, where $\delta(\Omega)$ is the Dirac delta function and the amplitude is given by

$$\Phi_0(z)=A_0\exp\left[-\Gamma(0)z+\frac{ik_{|\varphi|^2}}{2}\frac{A_0^2}{\Gamma(0)}(1-e^{-2\Gamma(0)z})\right], \tag{6}$$

where A_0 is the initial wave amplitude. In the derivation of this last expression we assumed that $k(0)=0$.

We shall analyze the evolution of this wave with an arbitrary initial perturbation $\theta(t)=\int \theta_\Omega \exp(i\Omega t)d\Omega$, assuming that

$$\varphi_\Omega(0)=A_0\delta(\Omega)+\theta_\Omega. \tag{7}$$

We shall postulate that the evolution-induced change in the perturbation is small compared with the amplitude (6) of the fundamental-frequency wave. The solution of the nonlinear equation can then be found in the form of a main solution and a correction:

$$\varphi_\Omega(z)=\Phi_0(z)\,\delta(\Omega)+\Phi_1(z,\Omega). \tag{8}$$

To first order in the small perturbation amplitude

$$\Phi_1(z,t)=\int \Phi_1(z,\Omega)\exp(-i\Omega t)\ll\Phi_0(z)$$

we can obtain coupled equations for the amplitudes of the Fourier representation $\Phi_1(z,\Omega)$:

$$\begin{aligned}\Phi_1'(z,\Omega)&=(-\Gamma_+ + ik_+)\Phi_1(z,\Omega)\\&\quad+ik_{|\varphi|^2}A_0^2e^{-2\Gamma_0 z}\left\{2\Phi_1(z,\Omega)\right.\\&\quad\left.+\exp\left[ik_{|\varphi|^2}\frac{A_0^2}{\Gamma_0}(1-e^{-2\Gamma_0 z})\right]\Phi_1^*(z,-\Omega)\right\},\end{aligned} \tag{9}$$

$$\begin{aligned}\Phi^{*\prime}_1(z,-\Omega)&=(-\Gamma_- - ik_-)\Phi_1^*(z,-\Omega)\\&\quad-ik_{|\varphi|^2}A_0^2e^{-2\Gamma_0 z}\left\{2\Phi_1^*(z,-\Omega)\right.\\&\quad\left.+\exp\left[-ik_{|\varphi|^2}\frac{A_0^2}{\Gamma_0}(1-e^{-2\Gamma_0 z})\right]\Phi_1(z,\Omega)\right\},\end{aligned} \tag{10}$$

where

$$\Gamma_+=\Gamma(\Omega)\quad \Gamma_-=\Gamma(-\Omega)\quad \Gamma_0=\Gamma(0)$$

$$k_+=k(\Omega)\quad k_-=k(-\Omega).$$

These equations give pairwise relationships between the amplitudes of the positive and negative frequencies. In the general case, the solution of the systems (9) and (10) may be expressed in the form

$$\Phi_1(z,\Omega)=F_+(z,\Omega)\,\theta(\Omega)+F_-(z,\Omega)\,\theta^*(-\Omega), \tag{11}$$

where $F_\pm(z,\Omega)$ are the solutions of this system for different initial conditions for positive and negative frequencies, specifically

$$F_+(z,\Omega)=\Phi_1(z,\Omega)\quad \text{for } \Phi_1(0,\Omega)=1,$$

$$\Phi_1^*(0,-\Omega)=0,$$

$$F_-(z,\Omega)=\Phi_1(z,\Omega)\quad \text{for } \Phi_1(0,\Omega)=0,$$

$$\Phi_1^*(0,-\Omega)=1.$$

The asymptotic form of the solution (11) in the nondissipative limit $\Gamma(\Omega)=0$ is easily found from the coupled system of equations and has the form

$$\begin{aligned}F_+(z,\Omega)&=\frac{\exp(ik_{|\varphi|^2}A_0^2z)}{\lambda_2-\lambda_1}[(-ik_+ +\lambda_2-ik_{|\varphi|^2}A_0^2)e^{\lambda_1 z}\\&\quad+(ik_+-\lambda_1+ik_{|\varphi|^2}A_0^2)e^{\lambda_2 z}],\\F_-(z,\Omega)&=\frac{\exp(ik_{|\varphi|^2}A_0^2z)}{\lambda_2-\lambda_1}(-ik_{|\varphi|^2}A_0^2e^{\lambda_1 z}\\&\quad+ik_{|\varphi|^2}A_0^2e^{\lambda_2 z}),\end{aligned} \tag{12}$$

where

$$\lambda_{1,2}=i\frac{k_+-k_-}{2}\pm\kappa,$$

$$\kappa^2=-\left(\frac{k_++k_-}{2}\right)^2-2k_{|\varphi|^2}A_0^2\frac{k_++k_-}{2}.$$

If we now assume that the initial perturbation is a δ-correlated random signal (white noise) for which

$$\langle\theta_\Omega\rangle=0,\quad \langle\theta_{\Omega1}\theta_{\Omega2}\rangle=0,\quad \langle\theta_{\Omega1}\theta^*_{\Omega2}\rangle=S\delta(\Omega_1-\Omega_2), \tag{13}$$

where $S=\text{const}$ is the spectral density of the white noise, then using the solution (11) we can find the spectral intensity of the output signal:

$$\begin{aligned}\langle\Phi_1(\Omega_1,z)\Phi_1^*(\Omega_2,z)\rangle&=S\delta(\Omega_1-\Omega_2)(|F_+(z,\Omega_1)|^2\\&\quad+|F_-(z,\Omega_1)|^2).\end{aligned} \tag{14}$$

We can show that in the general case ($\Gamma\neq0$) the solutions $F_\pm(z,\Omega)$ can be expressed in terms of hypergeometric functions. However, we merely note that in the following section we shall give the results of a numerical integration of a more general initial system. In the asymptotic limit $\Gamma=0$ we have

$$\begin{aligned}\langle\Phi_1(\Omega_1,z)\Phi_1^*(\Omega_2,z)\rangle&=S\delta(\Omega_1-\Omega_2)\\&\quad\times\left[1+\frac{2(k_{|\varphi|^2}A_0^2)^2}{\kappa^2}\sinh^2\kappa z\right].\end{aligned} \tag{15}$$

It can be seen that modulational instability occurs when $\kappa^2>0$. These formulas show that the spectral density of the amplified noise differs slightly from the expression for the frequency dependence of the amplified deterministic signal

$$|\Phi_1(z,\Omega)|^2=S_0|F_+(z,\Omega)+F_-(z,\Omega)|^2, \tag{16}$$

where S_0 is the initial intensity of the harmonic perturbation. At large distances this difference becomes more appreciable, as will become clear from the numerical calculations.

3. NUMERICAL SIMULATION OF SELF-MODULATION NOISE AMPLIFICATION

In order to take account of the initial noise perturbation of the signal wave and describe its nonlinear evolution, it is

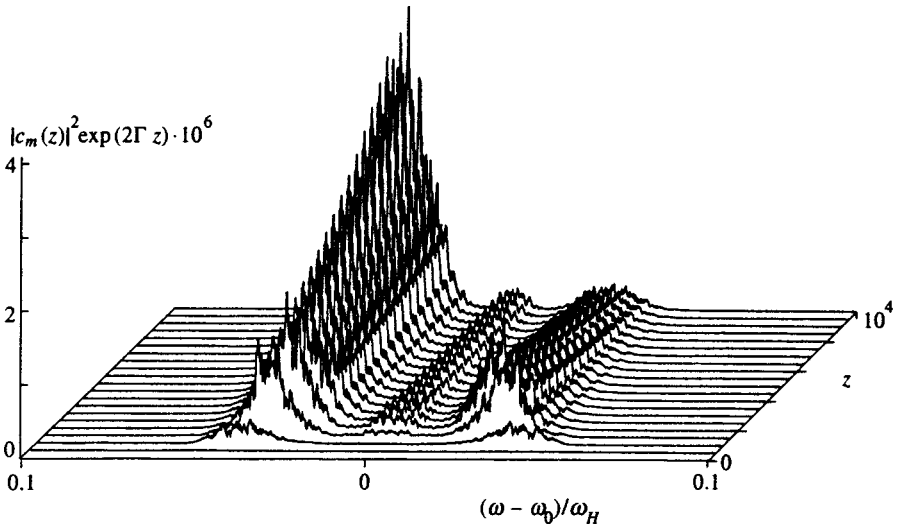


FIG. 1. Normalized intensity of noise amplified by self-modulation at initial signal wave powers $|C_0|^2=0.01$ and noise $|\epsilon^0_{(m\neq 1000)}|^2=9\times10^{-12}$ ($0\leqslant m\leqslant 2000$, $\Delta_\omega=10^{-4}$). Calculations using formula (18) were made for the spectrum of forward internal waves for the film parameters corresponding to the data given in Ref. 6: magnetization $4\pi M=1750$ G, line width $\Delta H=0.3$ Oe, film thickness $d=0.5$ μm, bias field $H=2500$ Oe, carrier frequency $\omega_0=2.4$ GHz. The frequencies in the figure are normalized to $\omega_H=\gamma(H-4\pi M)$ and the distance along the z axis is measured in film thicknesses d.

convenient to use a different approach, based on a spectral representation of the desired solution in the form of an expansion as a Fourier series:

$$\varphi(t,z)=\sum_{m=0}^{2N} c_m(z)\exp[i(m-N)\Delta_\omega t], \tag{17}$$

where $\Delta_\omega=2\pi/\omega$. In this case, the evolution of the amplitudes of the harmonics of the Fourier expansion is described by a normalized system of equations in the form

$$\frac{dc_l}{dz}=-\Gamma c_l-ik_lc_l-2i\sum_{m=0}^{2N} c_mS_{2N+m-l}, \tag{18}$$

where $l=0,\ \ldots,\ 2N$,

$$S_m=\sum_{n=2N-m}^{2N} c_nc^*_{2N-m+n}, \quad m=0,\ \ldots,\ 2N,$$

$$S_m=S^*_{4N-m}, \quad m=2N,\ \ldots,\ 4N,$$

$k_l=k(\Omega_l)$, $\Omega_l=(l-N)\Delta_\omega$. In Eq. (18) we use the normalization of the amplitude $[\Phi]=A_0\sqrt{k_{|\varphi^2|}}$. In this case, the initial condition for the deterministic perturbation scheme has the form

$$c_N(0)=C_0, \quad c_{N\pm l}(0)=\epsilon^0_{\pm 1}, \quad c_{m\neq N,N\pm l}(0)=0.$$

The noise perturbation is simulated under the initial conditions

$$c_N(0)=C_0, \quad c_{m\neq N}(0)=\epsilon^0_m\exp(i\beta^0_m),$$

where the amplitudes ϵ^0_m have a random normal distribution with a mean of zero and the phases β^0_m are uniformly distributed in the range $0<\beta^0_m<2\pi$. The spectral intensity of the noise is determined by the mean-square deviation in accordance with the formula $S_\omega=10\log\langle\epsilon^2_m\rangle$ dB. The Runge–Kutta method was used for a numerical solution of the system (18).

Figures 1–3 give results of a numerical simulation of the self-modulation effect in cases of noise and deterministic initial perturbations for parameters consistent with the experimental data presented in Ref. 6, in which the following experiments were carried out using several yttrium iron garnet (YIG) samples. The samples were YIG films 0.5–5 μm thick in which forward internal magnetostatic spin waves were excited using microstrip antennas 0.5–7 mm apart. The signal from the output antenna was fed to a spectrum analyzer. At low input powers the propagation was linear. Above a certain threshold power, 0.5–1 mW, discrete frequency satellites appeared in the spectrum, equidistant from the carrier frequency by frequencies of the order of 50–300 MHz. As

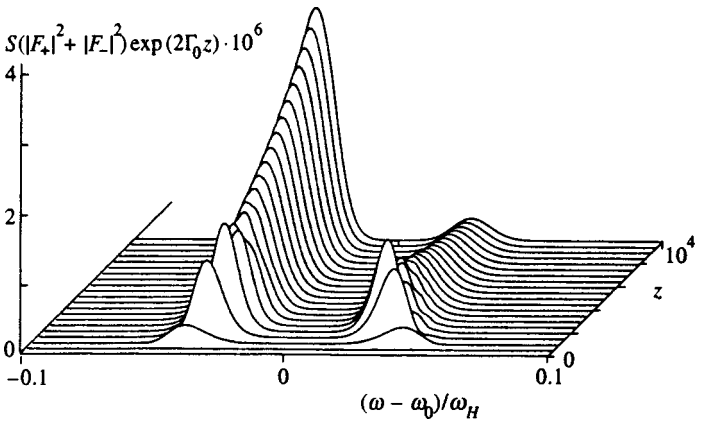


FIG. 2. Normalized intensity of self-modulation noise in accordance with formula (14) for $S=9\times10^{-12}$.

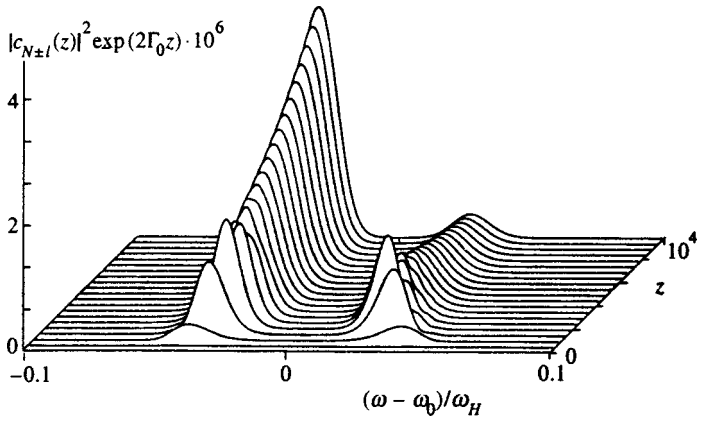


FIG. 3. Normalized intensity of self-modulation signal for a deterministic initial perturbation with $\epsilon^0_{\pm1}=3\times10^{-6}$.

the input signal power increased, the number of satellites increased and a broad noise-like peak formed from the low-frequency satellites. A high-frequency peak was not usually observed or was of substantially lower amplitude. Figures 1 and 2 show the evolution of the deterministic signal and noise calculated using Eqs. (18) and (14), respectively and Fig. 3 shows that calculated using formula (16). The frequency range of the self-modulational amplification of the noise agrees with the experimental data given above. It can be seen from the figures that the spectral distribution pattern of the intensity envelope of the output signal in the presence of initial white noise at short distances when the influence of nonlinear interaction is negligible, correlates with the deterministic model of frequency-independent instability evolution. Moreover, the calculations using the noise model of self-modulation obtained in the linear approximation (14) and using the deterministic model (16) almost agree. The calculations allow for the frequency dispersion of the damping which gives rise to asymmetry of the amplitude-frequency characteristic of the signal during the evolution of instability, as is usually observed experimentally. The relative amplitude of the right-hand frequency satellite, which undergoes strong damping, reaches a maximum and then decays. The relative amplitude of the left-hand satellite increases continuously over the selected propagation length and has a broad spectral width. According to these calculations, the frequencies of the peak satellites are separated from the carrier frequency by 80 MHz, which is consistent with the experimentally observed frequencies.

For higher intensities of the initial signal and longer propagation lengths the deterministic and noise models of modulational instability evolution predict different amplitude-frequency characteristics. This case is shown in Figs. 4 and 5, which give the results of calculations for $\Delta H=0.2$ Oe. The insets show the spectral intensities of the self-modulated signal for a short propagation length $z=250d$, where d is the film thickness. A comparison of the figures shows that the spectral distribution pattern of the normalized amplitudes of the signal intensity for the noise and deterministic models differs substantially at large distances.

In addition to the dispersion of the damping, the dispersion of the nonlinearity may also influence the asymmetry of the self-modulation spectrum in accordance with formula (4) for the growth rate. In subsequent experiments it would be extremely interesting to observe the dispersion of the spin wave nonlinearity and its influence on the behavior of the modulational instability, since numerical calculations allowing for the dispersion of the nonlinear frequency shift using the model formula

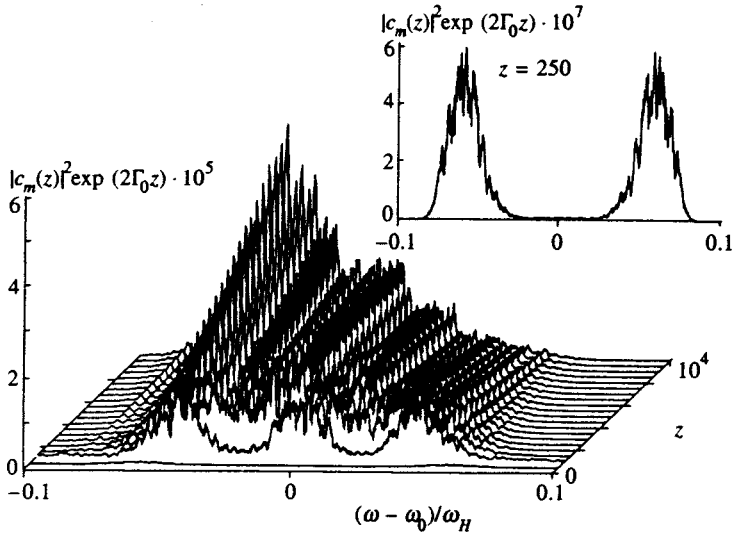


FIG. 4. Normalized intensity of self-modulation noise with reduced damping ($\Delta H=0.2$ Oe) and an increased signal wave amplitude $|C_0|^2=0.0144$.

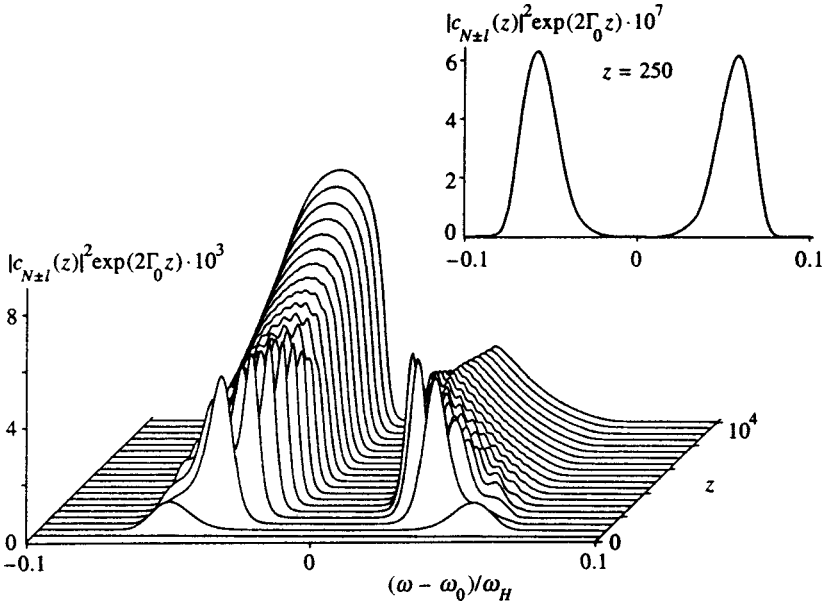


FIG. 5. Normalized intensity of self-modulation signal with a deterministic initial perturbation, with reduced damping and an increased signal wave amplitude, as in Fig. 4.

$$\frac{\partial\omega}{\partial|\varphi|^2}=\left(\frac{\partial\omega}{\partial|\varphi|^2}\right)_0+\alpha\Delta\omega,$$

have shown that the amplification of the right- and left-hand satellites in the noise model takes place at different growth rates (Fig. 6). The change in the dispersion associated with the nonzero third derivative does not cause any differences in the rates of amplification of the satellites. This is also deduced from a linear analysis of modulational instability, as can be seen from the formula (12).

4. CONCLUSIONS

To sum up, a simulation of the evolution of modulational instability as the amplification of a noise signal has shown that in the range of weak supercriticality, where the amplitude of the forming satellites is low, the resulting amplitude–frequency characteristic of the signal is broadly consistent with the model of frequency-independent amplification of a deterministic harmonic perturbation. A difference is observed for a long propagation length in a weakly dissipative medium when the relative amplitudes of the amplified noise at the satellite frequencies and the fundamental frequency become appreciable compared with the signal wave amplitude. Conditions of weak supercriticality are usually established in magnetic films because of the strong damping. Quite clearly, four-wave narrow-band, and broad-band self-modulational instabilities can coexist almost independently under these conditions. A similar situation was clearly observed in Ref. 6. At a minimum, the frequency interval for the self-modulational amplification of the noise calculated for the given film parameters agrees with the experimental data. The observed asymmetry of the amplification of neighboring self-modulation peaks can be attributed to the frequency dispersion of the nonlinear frequency shift and damping. This asymmetry is particularly large near the transmission boundaries of the magnetostatic waves in the film. Allowance for ordinary third-order dispersion does not produce this asymmetry.

For a further comparison between theory and experiment it would be useful to continue our investigation of modulational instability in YIG films. In particular, it would be interesting to study four-wave decay processes in thicker films (10–20 μm). In these films the wave dispersion is stronger which changes the conditions for the onset and buildup of

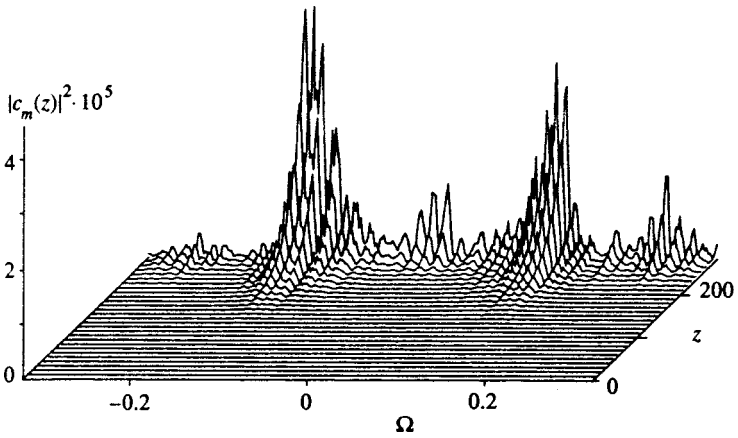


FIG. 6. Spectrum of self-modulation noise calculated using the system (18) with the dispersion of the nonlinearity $\omega_{|\varphi|^2}=2+60\Delta\omega$. A dispersion law in the form $k(\Omega)=\Omega+\Omega^2$ was analyzed. The initial signal wave power was $|C_0|^2=0.01$, and the noise was $|\epsilon^0_{m\neq 160}|^2=6.25\times 10^{-10}$ ($0\leqslant m\leqslant 320$, $\Delta_\omega=2\times 10^{-3}$).

instability. A comparison with the case of thin YIG films used experimentally will clarify the mechanism for the evolution of modulational instability in the beyond-threshold region and in particular, will reveal the more precise nature of the formation of a noise pedestal and the transition to chaos with allowance for the specific characteristics of magnetic films.

The authors are grateful to the Russian Fund for Fundamental Research for financial support (Grants Nos. 97-02-16183 and 96-02-16168) and also to the State Scientific Program on ''Surface Atomic Structures'' (Grant No. 97.3.16).

*)E-mail: nikitov@open.cplire.ru

[1] A. M. Mednikov, Fiz. Tverd. Tela (Leningrad) **23**, 242 (1981) [Sov. Phys. Solid State **23**, 136 (1981)].
[2] J. D. Adam, IEEE Trans. Magn. MAG–**16**, 1168 (1980).
[3] A. G. Temiryazev, Fiz. Tverd. Tela. (Leningrad) **29**, 313 (1987) [Sov. Phys. Solid State **29**, 179 (1987)].
[4] G. A. Melkov, and S. A. Sholom, Zh. Éksp. Teor. Fiz. **96**, 712 (1989) [Sov. Phys. JETP **69**, 403 (1989)].
[5] B. A. Kalinikos, N. G. Kovshikov, and A. N. Slavin, Pis'ma Zh. Tekh. Fiz. **10**, 936 (1984) [Sov. Tech. Phys. Lett. **10**, 392 (1984)].
[6] P. E. Zil'berman, S. A. Nikitov, and A. G. Temiryazev, JETP Lett. **42**, 110 (1985).
[7] Yu. V. Gulyaev, P. E. Zil'berman, S. A. Nikitov, and A. G. Temiryazev, Fiz. Tverd. Tela. (Leningrad) **28**, 2774 (1986) [Sov. Phys. Solid State **28**, 1553 (1986)].
[8] A. V. Vashkovskiĭ, V. I. Zubkov, É. G. Lokk, and S. A. Nikitov, Fiz. Tverd. Tela. (Leningrad) **30**, 827 (1988) [Sov. Phys. Solid State **30**, 475 (1988)].
[9] S. A. Nikitov, Su Jun, R. Marcelli, and P. De Gasperis, J. Magn. Magn. Mater. **145**, L6 (1995).
[10] Su Jun, S. A. Nikitov, R. Marcelli, and P. De Gasperis, J. Appl. Phys. **81**, 1341 (1987).
[11] G. T. Kazakov, A. V. Kozhevnikov, and Yu. A. Filimonov, Fiz. Tverd. Tela. (St. Petersburg) **39**, 330 (1997) [Phys. Solid State **39**, 288 (1997)].
[12] A. V. Lavrinenko, V. S. L'vov, G. A. Melkov, and V. B. Cherepanov, Zh. Éksp. Teor. Fiz. **81**, 1022 (1981) [Sov. Phys. JETP **54**, 542 (1981)].
[13] A. K. Zvezdin and V. F. Popkov, Zh. Éksp. Teor. Fiz. **84**, 606 (1983) [Sov. Phys. JETP **57**, 350 (1983)].
[14] B. A. Kalinikos, N. G. Kovshikov, and A. N. Slavin, Zh. Éksp. Teor. Fiz. **94**(2), 159 (1988) [Sov. Phys. JETP **67**, 303 (1988)].
[15] M. Chen, M. A. Tsankov, J. M. Nash, and C. E. Patton, Phys. Rev. Lett. **70**, 1707 (1993).
[16] M. Chen, M. A. Tsankov, J. M. Nash, and C. E. Patton, Phys. Rev. B **49**, 12773 (1994).
[17] A. D. Boardman, S. A. Nikitov, K. Xie, and H. Mehta, J. Magn. Magn. Mater. **145**, 357 (1995).
[18] V. S. L'vov, *Nonlinear Spin Waves* [in Russian], Nauka, Moscow (1987).
[19] M. Karlsson, J. Opt. Technol. **12**, 2071 (1995).
[20] A. N. Slavin and G. A. Dudko, J. Magn. Magn. Mater. **86**, 115 (1990).
[21] G. A. Dudko and Yu. A. Filimonov, Pis'ma Zh. Tekh. Fiz. **15**, 55 (1989) [Sov. Tech. Phys. Lett. **15**, 61 (1989)].
[22] N. N. Akhmedieva, V. M. Eleonskiĭ, and N. E. Kulagin, Zh. Éksp. Teor. Fiz. **89**, 1542 (1985) [Sov. Phys. JETP **62**, 894 (1985)].
[23] V. I. Karpman and E. M. Maslov, Zh. Éksp. Teor. Fiz. **73**, 537 (1977) [Sov. Phys. JETP **46**, 281 (1977)].
[24] A. D. Boardman, Q. Wang, S. A. Nikitov, J. Shen, W. Chen, D. Mills, and J. Bao, IEEE Trans. Magn. **MAG 30**, 14 (1994).

Translated by R. M. Durham

Coulomb effects in a ballistic one-channel $S-S-S$ device

D. A. Ivanov

L. D. Landau Institute for Theoretical Physics, 117940 Moscow, Russia;
12-127 M.I.T., Cambridge, MA 02139, USA

M. V. Feigel'man

L. D. Landau Institute for Theoretical Physics, 117940 Moscow, Russia
(Submitted 20 November 1997)
Zh. Éksp. Teor. Fiz. **114**, 640–653 (August 1998)

We develop a theory of Coulomb oscillations in superconducting devices in the limit of small charging energy $E_C \ll \Delta$. We consider a small superconducting grain with finite capacitance connected to two superconducting leads by nearly ballistic single-channel quantum point contacts. The temperature is assumed to be very low, so there are no single-particle excitations on the grain. Then the behavior of the system can be described in terms of the quantum mechanics of the superconducting phase on the island. The Josephson energy as a function of this phase has two minima that become degenerate when the phase difference on the leads equals to π, the tunneling amplitude between them being controlled by the gate voltage on the grain. We find the Josephson current and its low-frequency fluctuations, and predict their periodic dependence with period $2e$ on the induced charge $Q_x = CV_g$. © *1998 American Institute of Physics.* [S1063-7761(98)02008-3]

1. INTRODUCTION

Coulomb effects in several different types of three-terminal devices consisting of an island connected to external leads by two weak-link contacts, and capacitively coupled to an additional gate potential, have been extensively studied in the last few years. Systems with a normal metal island and leads were studied theoretically both in the tunnel-junction limit[1] and in the case of a quantum point contact with almost perfect transmission.[2] The theory of charge-parity effects and Coulomb modulation of the Josephson current was investigated in detail in Ref. 3. All of the above systems at present are realized experimentally.

Recently, it was shown to be possible to produce a quantum point contact between two superconductors via a normally conductive region made of two-dimensional electron gas;[4] smeared step-wise behavior of the critical current was observed, in qualitative agreement with predictions[5] for a superconductive quantum contact with a few conduction channels of high transmittivity. Observation of a nonsinusoidal current–phase relation in superconducting mechanically controllable break junctions has been reported in Ref. 6, again in agreement with Ref. 5.

Another interesting experimental achievement was reported in Ref. 7, where $S-N-S$ contact with a size comparable to the de Broglie wavelength in the N region made of BiPb was realized and nonmonotonic behavior of the critical current with the thickness of normal region was found. This remarkable development of technology suggests the feasibility of making a system of a small superconductive (SC) island connected to the superconductive leads by two quantum point contacts (QPC). In such a system, macroscopic quantum effects due to competition between Josephson coupling energy and Coulomb (charging) energy could be realized, together with quantization (due to the small number of conductive channels) of the Josephson critical current.

In the present paper we develop a theory for a limiting case of such a system, namely, two almost ballistic one-channel QPCs connecting a small SC island with two SC leads. We consider the limit of the characteristic charging energy much smaller than the superconducting gap, $E_C \ll \Delta$; therefore, Coulomb effects are small. We derive the dependence of the average Josephson current across the system and its fluctuations (noise power) as functions of the SC phase difference between the leads α, and of the electric gate potential V_g. Coulomb effects show up at phase differences α close to π, where the two lowest states are almost degenerate. We show that such a system realizes a tunable quantum two-level system (pseudospin 1/2) which may be useful for the realization of quantum computers (see, e.g., Refs. 8–11).

The paper is organized as follows. We start by considering a single QPC connecting a superconducting island to a single lead (Sec. 2). We find the oscillations of the effective capacitance of the island as a function of the gate potential (in some analogy with Matveev's results[2] for a normal QPC). Depending on the backscattering probability in the contact, it can be described either in the adiabatic or in the diabatic approximation. We find the condition for diabatic–adiabatic crossover. Then in Sec. 3 we formulate a simple model for a double-contact system in the adiabatic approximation. We replace the full many-body problem by a quantum-mechanical problem for the dynamics of the SC phase on the middle island. In Sec. 4 we calculate the average Josephson current through the system as a function of α and V_g, with particular emphasis on phase differences α close to π (where

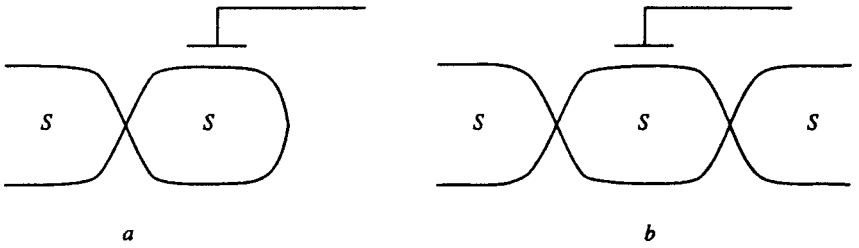


FIG. 1. a) Single QPC. The system consists of a SC grain connected to a SC lead via a QPC. A gate terminal is used to control the electric potential of the grain. b) Double-contact $S-S-S$ system. The second terminal is added to the single-QPC setup.

our effective two-level system is almost degenerate). Section 5 is devoted to the analysis of Josephson current noise; we calculate total intensity S_0 of the ''zero''-frequency noise (an analog of the noise calculated in Refs. 12–14 for a single superconductive QPC), as well as finite-frequency noise S_ω due to transitions between the two almost-degenerate levels. Finally, we present our conclusions in Sec. 6.

2. ADIABATIC–DIABATIC CROSSOVER IN A SINGLE QUANTUM POINT CONTACT

Consider a small superconducting island connected to an external superconducting lead by an one-channel, nearly ballistic quantum point contact.[5,15] The electric potential of the grain can be adjusted via a gate terminal (Fig. 1a). Following Ref. 5, we assume that the contact is much wider than the Fermi wavelength (so that transport through the constriction can be treated adiabatically), but much smaller than the coherence length $\xi_0 \equiv \hbar v_F/\pi\Delta$ (where v_F is the Fermi velocity, and Δ is the superconducting gap).

Our low-temperature assumption is that the average number of one-electron excitations on the island is much less than one. Then they cannot contribute to the total charge of the grain, and we may restrict our Coulomb blockade problem to the evolution of the superconducting phase only. The low-temperature condition is then $T<\Delta/\ln(V\nu(0)\Delta)$, where V is the volume of the grain and $\nu(0)$ is the density of electron states at the Fermi level.

We neglect phase fluctuations in the bulk of the island and describe the whole island by a single superconducting phase χ. At a fixed value of the phase on the island, the spectrum of the junction consists of the two Andreev states localized on the junction and the continuum spectrum above the gap[15] Δ (Fig. 2). The energies of the Andreev states lie below the gap:

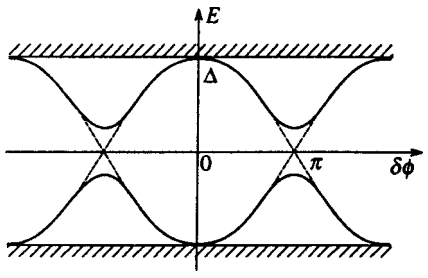


FIG. 2. Single-contact energy spectrum. The spectrum consists of the continuum of delocalized states and the two Andreev (subgap) states. Dashed lines denote Andreev states in the absence of backscattering (diabatic terms). Solid lines are the states split by backscattering (adiabatic terms).

$$E(\chi) = \pm\Delta\sqrt{1 - t\,\sin^2(\chi/2)}, \tag{1}$$

where χ is the phase difference at the contact and t is the transmission coefficient.

At $t=1$, the spectrum of Andreev states (1) has a level crossing point at $\chi=\pi$. At this point, the left and right Andreev states have equal energies, but in the absence of backscattering ($t=1$), transitions between them are impossible. Therefore, we expect that an ideal ballistic contact cannot adiabatically follow the ground state as the phase χ changes, but remains in the same left or right Andreev state as it passes the level-crossing point $\chi=\pi$. We borrow the terminology from the theory of atomic collisions[16] and call the (crossing) Andreev levels at $t=1$ diabatic terms (dashed lines in Fig. 2), and the split levels—adiabatic terms (solid lines in Fig. 2). Instead of a transmission coefficient t, it will be more convenient to speak of the reflection coefficient $r=1-t$. At $r=0$, the contact is described by diabatic terms. As r increases, transitions occur between the terms, and at sufficiently large r the system will mostly adiabatically follow the split Andreev levels. In this section we study the adiabatic–diabatic crossover, and find the crossover scale for the reflection coefficient r.

We assume that the reflection probability $r \ll 1$ (almost perfect transmission) and that the charging energy $E_C \ll \Delta$ (the charging energy is defined by $E_C=(2e)^2/C$). The latter assumption appears natural, because as in tunnel junctions[17] we expect that the capacitance C of the grain has an additional contribution from the capacitance of the point contact. This capacitance is of order Δ/e^2. A more detailed discussion of this phenomenon will be given elsewhere. For now, we just mention that this contribution to the capacitance leads to the inequality $E_C \leqslant \Delta$.

To probe the degree of adiabaticity, we study the periodic dependence of the ground state energy E_0 on the gate voltage. Because of the weakness of charging effects, this dependence will be sinusoidal:

$$E_0(V_g) = \varepsilon\,\cos(2\pi N) \tag{2}$$

(where $N=V_gC/2e$ is the dimensionless voltage), and we are interested in the amplitude ε of these oscillations. The physical origin of this periodicity is oscillations of the induced charge on the grain; this follows immediately from the relation

$$\delta Q = \frac{C}{2e}\,\frac{\partial E_0}{\partial N}. \tag{3}$$

There is a simple physical explanation of the sinusoidal dependence (2). The ground-state energy modulation is de-

termined by phase-slip processes in the contact. Such processes are phase tunneling events with phase changing by $\pm 2\pi$. While the magnitudes of the clockwise and counterclockwise tunneling amplitudes are the same, their phases are $\pm 2\pi N$. This results in the expression (2). Higher-order tunneling processes would give rise to higher-order harmonics in the periodic N-dependence. This argument shows that the amplitude of oscillations ε coincides with the phase-tunneling amplitude, and therefore provides a good measure of adiabaticity in the phase dynamics.

Assuming $E_C \ll \Delta$, we can describe the contact by the dynamics of the phase on the grain, and thus reduce the problem to single-particle quantum mechanics. Since we restrict our attention to low-lying excitations, it is only necessary to include the two Andreev levels on the junction. The potential term is the Josephson energy of the Andreev levels, and the kinetic term is the charging energy. After a simple computation of the backscattering matrix elements (the off-diagonal entries in the potential term), we arrive at the Hamiltonian:

$$H = H(\chi) + \frac{1}{2} E_C (\pi_\chi - N)^2, \tag{4}$$

where

$$H(\chi) = \Delta \begin{pmatrix} -\cos\frac{\chi}{2} & r^{1/2}\sin\frac{\chi}{2} \\ r^{1/2}\sin\frac{\chi}{2} & \cos\frac{\chi}{2} \end{pmatrix}. \tag{5}$$

Here χ is the phase difference across the contact, and r is the reflection coefficient. Obviously, the eigenvalues of $H(\chi)$ reproduce the result (1). The number of Cooper pairs at the grain π_χ is the momentum conjugate to χ, $[\chi, \pi_\chi] = i$. Notice that χ takes values on the circle $\chi = \chi + 2\pi$, and accordingly π_χ is quantized to take integer values. We can also write $\pi_\chi = -i\partial/\partial\chi$.

This Hamiltonian loses its validity at the top of the upper band at $\chi = 2\pi n$, where the upper Andreev state mixes with the continuous spectrum (Fig. 2). However, the probability of the phase χ reaching the top of the upper band of $H(\chi)$ is exponentially small at $E_C \ll \Delta$ (smaller than the tunneling probability). The adiabatic–diabatic crossover is determined by the properties of the system near the minimal-gap point $\chi = \pi$. We can therefore neglect transitions to the continuous spectrum at $\chi = 2\pi n$. At the same time, we must disregard tunneling processes via the top of the upper Andrees band (next-nearest-neighbor tunneling), which are present in the Hamiltonian (4)–(5), but not in the original system. Nearest-neighbor tunneling is a feature of our model, and is beyond the precision of our approximation.

There are two opposite limits of the problem: small and ''strong'' reflection.

At zero reflection, the Hamiltonian splits into lower and upper components. Within each component the potential is periodic with period 4π. As explained above, we must neglect next-nearest-neighbor tunneling via the top of the bands. Therefore, the potential minima of $H(\chi)$ are disconnected and cannot tunnel to each other ($\varepsilon = 0$).

The opposite limit is the case of ''large'' reflection (the precise meaning of ''strong reflection'' consistent with $r \ll 1$ will be clarified below). In this limit, a gap opens in the spectrum of Andreev states, and the system adiabatically follows the lower state. We can replace the two-level Hamiltonian $H(\chi)$ by its lowest eigenvalue and arrive to the quantum-mechanical problem of a particle in a periodic potential. The semiclassical limit of this problem is solved in Ref. 18. In our notation, the result is

$$\varepsilon_{\rm ad} = \text{const}\ \sqrt{E_C \Delta}\ \exp(-S_{\rm cl}), \tag{6}$$

where

$$S_{\rm cl} = B_1 \sqrt{\frac{\Delta}{E_C}} - \frac{1}{4}\ln\frac{\Delta}{E_C} + O(1) \tag{7}$$

is the classical action connecting two nearest minima (or more precisely, the two turning points). The numerical constant B_1 is of order unity (at $r \to 0$, $B_1 = 4.69 + 1.41 r \ln r + \ldots$).

To study how the adiabaticity is destroyed, it is useful to introduce the dimensionless ''coherence factor'' $f(r)$ defined by

$$\varepsilon = f(r)\varepsilon_{\rm ad}, \tag{8}$$

where $\varepsilon_{\rm ad}$ is the amplitude of oscillation of the ground-state energy derived in the adiabatic approximation (with only the lowest Andreev state included). We see that $f(0) = 0$ and $f(r \gg r_{\rm ad}) = 1$. The crossover scale $r_{\rm ad}$ can be derived by computing the corrections to $f(r)$ in these two limits.

We first consider the limit of weak backscattering ($r \ll r_{\rm ad}$). In this limit, we take the wavefunction to be the ground state of the Hamiltonian with zero r (at a given wavevector N), and then compute the first-order correction in $r^{1/2}$ to the energy. The wavefunction is of ''tight-binding'' type, and is generated by the ''ground-state'' wavefunctions Ψ_i localized in the potential minima (diabatic terms). The components of the two-dimensional vectors Ψ_i alternate:

$$\Psi_i = \begin{pmatrix} \Psi_i(\chi) \\ 0 \end{pmatrix}, \quad \Psi_{i+1} = \begin{pmatrix} 0 \\ \Psi_{i+1}(\chi) \end{pmatrix}. \tag{9}$$

We then find

$$\varepsilon = 2\langle \Psi_i | H_{12}(\chi) | \Psi_{i+1} \rangle = 2 r^{1/2} \Delta \int d\chi \Psi_i^*(\chi) \Psi_{i+1}(\chi) \sin\frac{\chi}{2} \tag{10}$$

(we assume the wavefunctions Ψ_i to be normalized). It is important to note that Ψ_i and Ψ_{i+1} are wavefunctions for different potentials ($-\Delta_0 \cos(\chi/2)$ and $\Delta_0 \cos(\chi/2)$); the overlap integral (10) has a saddle point at the minimal-gap point $\chi = \pi$, and it reduces the effective region of integration to $|\chi - \pi| \leqslant (E_C/\Delta)^{1/4}$. The normalization of the semiclassical tail of the wavefunctions $\Psi_i(\chi)$ yields

$$\Psi(\chi = \pi) = \exp(-S_{\rm cl}(\chi = \pi)) \tag{11}$$

(up to a numerical factor independent of E_C/Δ). We thus obtain

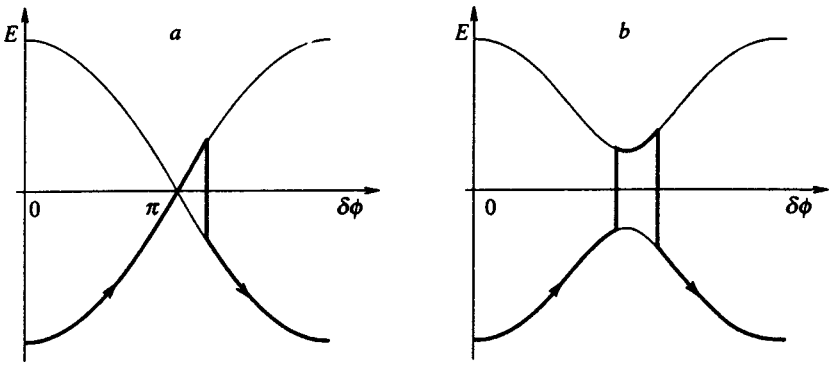


FIG. 3. Tunneling paths in the diabatic a) and adiabatic b) limits. These diagrams represent the lowest-order corrections to the phase-tunneling amplitudes in the diabatic and adiabatic limits, respectively.

$$\varepsilon \sim r^{1/2}\Delta\left(\frac{E_C}{\Delta}\right)^{1/4} \exp(-S_{\rm cl}), \tag{12}$$

i.e., in terms of the ''coherence factor'' $f(r)$,

$$f(r)\sim r^{1/2}\left(\frac{\Delta}{E_C}\right)^{1/4}. \tag{13}$$

Physically, meaning of the integral (10) is the sum over all paths shown in Fig. 3a.

The above calculation shows that the crossover scale to adiabatic behavior is

$$r_{\rm ad}\sim\left(\frac{E_C}{\Delta}\right)^{1/2}. \tag{14}$$

In fact, we neglected the effect of change in the classical action $S_{\rm cl}$ due to the gap opening; this effect is estimated to be of order

$$\delta S_{\rm cl}\sim \sqrt{\frac{\Delta}{E_C}}\, r\, \ln r, \tag{15}$$

i.e., it is a higher-order effect than the change in $f(r)$ proportional to $r^{1/2}$. Notice that the characteristic scale of this change in the classical action is again $r_{\rm ad}\sim\sqrt{E_C/\Delta}$ (corresponding to $\delta S_{\rm cl}\sim 1$).

We can alternatively find the crossover scale $r_{\rm ad}$ by computing the lowest order correction to the ''coherence factor'' $f(r)$ in the adiabatic limit. In this limit the Hamiltonian (4), (5) can be rewritten in adiabatic terms (for simplicity the voltage N is introduced into the boundary condition $\Psi(\chi+2\pi)=e^{2i\pi N}\Psi(\chi)$ by a gauge transformation) as

$$H=-\frac{E_C}{2}\left(\frac{\partial}{\partial\chi}\right)^2+D(\chi) - \frac{E_C}{2}\left[G(\chi)\,\frac{\partial}{\partial\chi}+\frac{\partial}{\partial\chi}\,G(\chi)\right]-\frac{E_C}{2}\,G^2(\chi), \tag{16}$$

where

$$D(\chi)=\begin{pmatrix} E_1(\chi) & 0\\ 0 & E_2(\chi)\end{pmatrix} \tag{17}$$

is the diagonalized form of the matrix (5),

$$G(\chi)=\begin{pmatrix} 0 & g(\chi)\\ -g(\chi) & 0\end{pmatrix},\quad g(\chi)=\langle 0|\,\frac{\partial}{\partial\chi}\,|1\rangle, \tag{18}$$

and $|0\rangle$ and $|1\rangle$ are the eigenvectors of the matrix (5). The last term in the Hamiltonian (16) can be shown to yield smaller corrections than the first-order term in $G(\chi)$. A careful second-order perturbation calculation in $g(\chi)$ yields

$$1-f(r)\sim\int_{\chi_1<\chi_2}\exp\{S_1(\chi_1,\chi_2)-S_2(\chi_1,\chi_2)\} \times g(\chi_1)g(\chi_2)d\chi_1 d\chi_2, \tag{19}$$

where $S_{1,2}(\chi_1,\chi_2)$ are the classical actions along the lower and upper adiabatic branches between the points χ_1 and χ_2. This integral corresponds to summation over all tunneling paths shown in Fig. 3(b). The function $g(\chi)$ for the given matrix $H(\chi)$ is a Lorentzian peak at $\chi=\pi$ of height $r^{-1/2}$ and width $r^{1/2}$. Putting everything together, the integral (19) is calculated to be

$$1-f(r)\sim\frac{1}{r}\,\sqrt{\frac{E_C}{\Delta}}. \tag{20}$$

This asymptotic behavior agrees with the crossover scale (14) found previously.

To summarize the results of this section, the characteristic scale for adiabatic–diabatic crossover in a nearly-ballistic single contact is found to be $r_{\rm ad}\sim\sqrt{E_C/\Delta}$. The phase tunneling amplitude is proportional to the gate-voltage modulation of the effective capacitance of the island, and thus can be directly measured. At low reflection coefficients, these oscillations are proportional to $\sqrt{r}$, as in the normal one-channel QPC.[2]

3. ADIABATIC APPROXIMATION OF A DOUBLE-JUNCTION SYSTEM

We now turn to the case of a double-junction system (Fig. 1b). As before, we assume that the reflection probabilities in both contacts are small, $r_i\ll 1$, that the charging energy $E_C\ll\Delta$, and that the temperature is sufficiently low to preclude single-electron excitations on the grain. To adjust the electrostatic potential of the grain we again use a gate terminal; $N=V_gC/2e$ denotes the dimensionless gate voltage, as before.

For the moment, to simplify the discussion we assume that the reflection coefficients in the contacts are greater than the crossover scale $r_{\rm ad}$ found in the previous section; we can therefore consider only the lower adiabatic branch of the

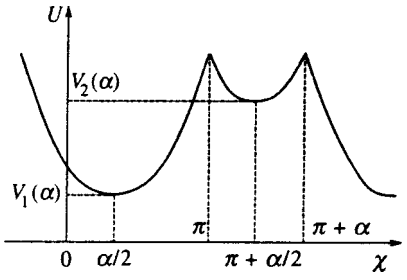


FIG. 4. Potential $U(\chi)$. At $\alpha \neq 0$ it has two minima. Finite backscattering in the contacts smoothes the summits of the potential, but leaves the bottom of the wells unchanged.

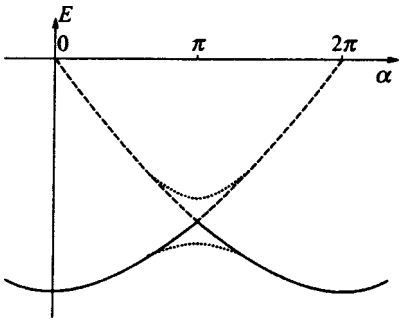


FIG. 5. Classical minimum of the potential $U(\chi)$ as a function of the external phase difference α. Dotted line shows the quantum gap opened by the Coulomb term.

Andreev states. In fact, the results can be extended further to the case $r_i < r_{\rm ad}$ by using appropriate ''coherence factors'' $f(r)$, similar to those in the previous section.

We set the superconducting phase on one of the leads to zero; the phase on the other lead α is assumed to be fixed externally. Then the total Josephson energy of the two contacts is (Fig. 4):

$$U(\chi) = U_1(\chi) + U_2(\alpha - \chi), \tag{21}$$

where

$$U_i(\delta\phi) = -\Delta\sqrt{1 - t_i \sin^2(\delta\phi/2)} \tag{22}$$

are the lower adiabatic Andreev terms in the two junctions.

At $t_1 = t_2 = 1$, the potential $U(\chi)$ obviously has two minima—at $\chi = \alpha/2$ and at $\chi = \alpha/2 + \pi$—and sharp peaks at $\chi = \pi$ and $\chi = \pi + \alpha$ (Fig. 4). At small nonzero r_i, gaps open at the crossing points of Andreev levels, which smoothes the peaks of $U(\chi)$. Still, the bottom of the potential remains essentially unchanged.

The adiabatic Hamiltonian for the double junction becomes

$$H(\alpha, N) = U(\chi) + U(\alpha - \chi) + \frac{1}{2} E_C \left(-i \frac{\partial}{\partial \chi} - N \right)^2. \tag{23}$$

The potential term of the Hamiltonian is the sum of Josephson energies of the contacts, and the kinetic term is the Coulomb energy of the charge at the grain.

4. JOSEPHSON CURRENT

The condition $E_C \ll \Delta$ enables us to treat the Coulomb term in the Hamiltonian perturbatively. First, neglecting the Coulomb term, we obtain a classical system defined on the circle $\chi \in (0, 2\pi)$ in the potential (21) with two minima. The energies of the minima are $V_1(\alpha) = -2\Delta|\cos(\alpha/4)|$ and $V_2(\alpha) = -2\Delta|\sin(\alpha/4)|$ (see Fig. 4). To very high accuracy, we can neglect backscattering in determining the minima, except near $\alpha = 0$. Since all Coulomb effects occur near the resonance point $\alpha = \pi$, this approximation is justified. At zero temperature, our classical system prefers the lowest of the minima. Thus the energy of the $S-S-S$ system in the absence of the Coulomb term is given by

$$E(\alpha) = -2\Delta \cos(\alpha/4) \quad \text{for} \quad -\pi < \alpha < \pi \tag{24}$$

(see Fig. 5). Differentiating this energy with respect to the phase α gives the Josephson current

$$I(\alpha) = 2e \frac{\partial E(\alpha)}{\partial \alpha} = \Delta \sin\frac{\alpha}{4} \quad \text{for} \quad -\pi < \alpha < \pi \tag{25}$$

(Fig. 6). Notice that the current has large jumps at the points of level crossing $\alpha = \pi + 2\pi n$. Qualitatively this picture is very similar to the case of a single $S-S$ ballistic junction, but the shape of the current–phase dependence $I(\alpha)$ is different.

If we assume a nonzero temperature $T \ll \Delta$, the occupation of the upper minimum is exponentially small except in the vicinity of the level-crossing point $|\alpha - \pi| \sim T/\Delta$. Thus, the effect of the temperature is to smear the singularity in $I(\alpha)$ at $\alpha = \pi$.

Another source of level mixing near the singular point $\alpha = \pi$ is quantum fluctuations, i.e., fluctuations arising from the kinetic term in the Hamiltonian (23). They result in nonzero tunnelling amplitudes through the two potential barriers between the potential minima. Due to the shift in the ''angular momentum'' by N, the wave functions in the two potential wells acquire an additional factor $\exp(iN\chi)$. This results in the relative phase of the two tunneling amplitudes differing by $2\pi N$. The net tunneling amplitude (defining the level splitting) can be written

$$H_{12}(N) \equiv \Delta\gamma(N) = \Delta(\gamma_1 e^{i\pi N} + \gamma_2^{-i\pi N}), \tag{26}$$

where γ_1 and γ_2 are the two amplitudes of phase tunneling in the two different directions (i.e., of phase slip processes in the two different contacts). Below we assume that these am-

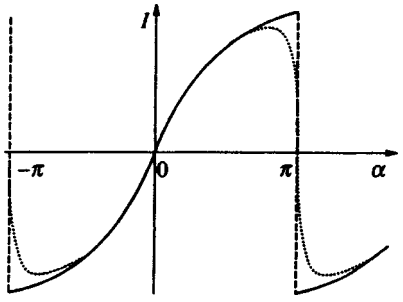


FIG. 6. Josephson current as a function of the external phase difference α. Dotted line shows smearing of the singularity due to the Coulomb term.

plitudes are computed at the level-crossing point $\alpha=\pi$, where they are responsible for level splitting.

The amplitudes γ_1 and γ_2 have the asymptotic behavior derived in the previous section (except for numerical factors). When the backscattering in the contacts is such that $r \gg r_{\rm ad}$, they can be found in the semiclassical approximation:

$$\gamma_{1,2}\sim\left(\frac{E_C}{\Delta}\right)^{1/4}\exp\left(-B_2\sqrt{\frac{\Delta}{E_C}}\right)\ll 1, \tag{27}$$

where $B_2\sim 1$ is determined by the classical action connecting the two potential minima (at $r\ll 1$, $B_2\simeq 1.45+2.20r\ln r+...$). At $r\ll r_{\rm ad}$, the tunneling amplitudes are

$$\gamma_{1,2}\sim r^{1/2}\exp\left(-B_2\sqrt{\frac{\Delta}{E_C}}\right). \tag{28}$$

For the best observation of Coulomb oscillations, γ_1 and γ_2 must be of the same order, but not too small. In the ideal case $\gamma_1=\gamma_2=\gamma$, the total amplitude is

$$\gamma(N)=2\gamma\cos(\pi N). \tag{29}$$

Although the periodic dependence (29) has a period of $4e$ as function of the ''external charge'' $Q_x=CV_g\equiv 2eN$, the Josephson current and its fluctuations only depend on $|\gamma(N)|^2$ (cf. Eqs. (32) and (34) below), and their period is $2e$, as expected.[3]

The characteristic scale for the r-dependence of B_2 is $\delta r\sim\sqrt{E_C/\Delta}$, so for γ_1 and γ_2 to be of the same order, the transparencies of the two contacts must differ by no more than $|r_1-r_2|\leqslant\sqrt{E_C/\Delta}$.

Here we should comment on the difference between our result (26)–(28) and the normal two-channel system discussed in Ref. 2. In the normal system, the two tunneling amplitudes multiply, and the net ground-state energy oscillations are proportional to $r\ln r$ at small r. In the superconducting system, the external leads have different superconducting phases, and the tunneling in the two contacts occurs at different values of the phase on the grain. Therefore, the tunneling amplitudes add with certain phase factors, and yield the asymptotic behavior $\sqrt{r}$ at $r\to 0$. In fact, the oscillations in the superconducting system will be proportional to r (as in the normal system[2]) in a different limit—at a phase difference $\alpha=0$, where the potential $U(\chi)$ has a single minimum and a single barrier.

The hybridized energy levels in the vicinity of $\alpha=\pi$ are given by the eigenvalues of the 2×2 Hamiltonian

$$H(\alpha,N)=\begin{pmatrix} V_1(\alpha) & H_{12}(N)\\ H_{12}(N) & V_2(\alpha)\end{pmatrix}. \tag{30}$$

Diagonalization yields the two energy levels:

$$E_{1,2}(\alpha,N)=-\Delta\left[\left|\sin\frac{\alpha}{4}\right|+\left|\cos\frac{\alpha}{4}\right| \pm\sqrt{\left(\left|\sin\frac{\alpha}{4}\right|-\left|\cos\frac{\alpha}{4}\right|\right)^2+\gamma^2(N)}\right]. \tag{31}$$

The off-diagonal matrix elements of the Hamiltonian open a gap at the level-crossing point $\alpha=\pi$ (Fig. 5). This gap depends periodically on the gate voltage V_g, and these oscillations comprise the Coulomb effects in the $S-S-S$ junction.

We can obtain the Josephson current by differentiating the energy levels with respect to the phase α. The gap results in smearing the singularity in $I(\alpha)$, even at zero temperature (Fig. 6):

$$I(\alpha)=\frac{\Delta}{\sqrt{2}}\sin\left(\frac{\alpha-\pi}{4}\right)\times\left[1-\frac{\cos\left(\frac{\alpha-\pi}{4}\right)}{\sqrt{\sin\left(\frac{\alpha-\pi}{4}\right)+\frac{1}{2}\gamma^2(N)}}\right] \quad \text{for } \alpha\sim\pi. \tag{32}$$

The width of the crossover at $\alpha=\pi$ depends periodically on V_g: $|\alpha-\pi|\sim|\gamma(N)|$.

In the above discussion we neglected excited oscillator states. The interlevel spacing for the excitations in the potential wells is of order $\sqrt{\Delta E_C}\gg\Delta\gamma$. Therefore, Coulomb effects have a much smaller energy scale and the excited states do not participate in mixing the ground states of the two potential wells.

At nonzero temperature, these Coulomb effects compete with temperature-induced smearing, so that the width of the singularity at $\alpha=\pi$ is given at nonzero temperature $T\ll\Delta$ by $|\alpha-\pi|\sim\max(\gamma(N),T/\Delta)$. In order for Coulomb effects to dominate thermal fluctuations, we must therefore have $T\leqslant\gamma\Delta$.

It is instructive to compare this picture with the case of a multi-channel $S-S-S$ tunnel junction (in contrast to the results of Ref. 3, note that we consider the opposite limit, with $\Delta\gg E_C$). If we develop a similar theory for tunnel Josephson junctions, we find that the potentials (21) and (22) are both sinusoidal, and therefore the total potential (21) has only one minimum (versus two in the nearly ballistic system). In the tunnel $S-S-S$ system, the current–phase relation $I(\alpha)$ is smeared at $\alpha=\pi$ due to the difference between the critical currents of the two Josephson junctions. Coulomb effects compete with this smearing, and in order to prevail, the charging energy E_C must be greater than the difference of the critical currents. In the tunnel system, the corresponding splitting γ is linear in E_C, while in the nearly ballistic system it is exponentially small. Otherwise, Coulomb oscillations in $I(\alpha)$ will appear similar in these two cases.

To summarize this section, we observed that the Coulomb effects in the one-channel $S-S-S$ junction smears the singularity in the Josephson current $I(\alpha)$ at the critical value $\alpha=(2n+1)\pi$. This smearing depends periodically on the potential of the grain with period $2e/C$, and is exponentially small in the adiabatic parameter $E_C/\Delta\ll 1$. The smearing is the result of mixing the two states in the potential minima of the Josephson energy.

5. FLUCTUATIONS OF THE JOSEPHSON CURRENT

In this section we compute the low-frequency spectrum of the fluctuations of the Josephson current in our model. We

shall be interested in frequencies much less than the oscillator energy scale $\sqrt{\Delta E_C}$, so we consider only transitions between the eigenstates of the reduced ground-state Hamiltonian (30). We also assume that the temperature is lower than $\sqrt{\Delta E_C}$; we can then disregard excited oscillator states and internal noise in the contacts (discussed in Refs. 12–14, 19). Obviously, under these assumptions we can observe current fluctuations only in the immediate neighborhood of the resonance point $\alpha = \pm\pi$, where the energies (31) of the two low-lying states are close to each other.

We expect to observe two peaks in the noise spectrum—one at zero frequency (due to thermal excitations above the ground state), and the other at the transition frequency $|E_1 - E_2|$ (due to off-diagonal matrix elements of the current operator). In this section we compute the total weights of these peaks and postpone discussion of their width (determined by dissipative processes).

Consider first the zero-frequency peak. In our approximation it is just the thermal noise of a two-level system. In the vicinity of the resonance point $\alpha = \pi$, we can linearize the spectrum $V_{1,2}(\alpha)$ and make the approximation that one of the two states carries the current $I(\alpha,N)$, and the other $-I(\alpha,N)$. The spectral weight of the noise is then given by a simple formula:

$$S_0(\alpha,N,T) \equiv \langle I^2\rangle - \langle I\rangle^2 = \frac{I^2(\alpha,N)}{\cosh^2 \dfrac{E_1 - E_2}{2T}}. \tag{33}$$

Substituting $I(\alpha,N)$ and $E_{1,2}(\alpha,N)$ from the previous section, we obtain the noise intensity near resonance:

$$S_0(\alpha,N,T) = \frac{\Delta^2}{2} \frac{\left(\dfrac{\alpha-\pi}{2\sqrt{2}}\right)^2}{\left(\dfrac{\alpha-\pi}{2\sqrt{2}}\right)^2 + \gamma^2(N)} \times \cosh^{-2}\left(\frac{\Delta}{T}\sqrt{\left(\frac{\alpha-\pi}{2\sqrt{2}}\right)^2 + \gamma^2(N)}\right). \tag{34}$$

For the effect of the Coulomb interaction to be observable, the temperature must be smaller than the Coulomb gap: $T \leqslant \gamma\Delta$. At constant T and N, the noise decreases exponentially as α moves away from its critical value $\alpha = \pi$, and at $\alpha = \pi$ the noise is suppressed in the interval $|\alpha - \pi| < \gamma(N)$ (Fig. 7). The interplay between these two factors results in a strong dependence of the peak noise on the potential of the grain. The peak value of the noise $\max_\alpha S(\alpha,N,T)$ is plotted against N in Fig. 8. Most favorable is the case of identical contacts, where $\gamma_1 = \gamma_2 = \gamma$, and therefore $\gamma(N) = 2\gamma \times \cos(\pi N)$. In this case, $\cos(\pi N) \ll T/\gamma\Delta$ (small gap limit) the noise takes its maximal value $S \approx \Delta^2/2$. In the opposite limit of large gap ($\cos(\pi N) \gg T/\gamma\Delta$), the noise decreases exponentially:

$$S \approx \Delta^2\left[\frac{T}{\Delta\gamma|\cos\ \pi N|}\exp\left(-4\frac{\Delta\gamma|\cos\ \pi N|}{T}\right)\right].$$

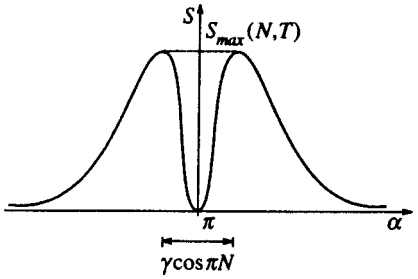


FIG. 7. Zero-frequency noise as a function of phase α. It decays exponentially for α far from the resonance point $\alpha = \pi$. Right at the very resonance point, the noise is suppressed, because both states carry nearly zero Josephson current.

The noise has a sharp peak at the resonance point $\cos \pi N = 0$, where two levels on the grain with different electron numbers have equal energies.

Now we turn to the noise peak at the interlevel frequency $\omega = |E_1 - E_2|$. Since ω can now be large compared to T, one needs to discriminate between different kinds of frequency-dependent correlation functions, which can be measured as a noise intensity in different experimental situations;[21] here, by noise we mean the Fourier spectrum of the time-symmetric current-current correlation function. In our approximation of a two-level system, such noise is temperature independent, and its weight is determined purely by the off-diagonal matrix element:

$$S_\omega = \frac{1}{2}\,|\langle 1|I|2\rangle|^2. \tag{35}$$

A straightforward computation for the Hamiltonian (30) and $I = 2e(\partial H/\partial\alpha)$ yields (in the vicinity of $\alpha = \pi$)

$$\langle 1|I|2\rangle = \frac{\Delta^2\gamma(N)}{\omega}\left(\cos\frac{\alpha}{4} + \sin\frac{\alpha}{4}\right) \tag{36}$$

and

$$S_\omega(\alpha,N) = \Delta^2\left(\frac{\Delta\gamma(N)}{\omega}\right)^2 \cos^2\frac{\alpha-\pi}{4}. \tag{37}$$

This result contrasts with the corresponding noise intensity in the single quantum point contact (found in Refs. 12, 13, and 19). In the single quantum point contact, the corresponding noise intensity S_ω is temperature-dependent, because that system has four possible states (or, alternatively, two fermion levels). In the case of the double junction, the

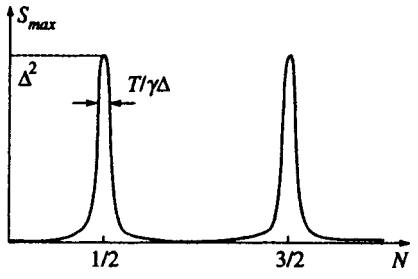


FIG. 8. Maximum value of the noise versus the potential of the grain. The period of the peaks corresponds to the period $2e$ of the induced charge $Q = CV_g$. The width of the peaks depends on the capacitance of the grain.

system has only two states differing by the phase on the grain, and the quantum fluctuations S_ω become temperature-independent.

6. CONCLUSIONS

We have developed a theory of Coulomb oscillations of the Josephson current and its noise power via the $S-S-S$ system with nearly ballistic quantum point contacts. The period of Coulomb oscillations as a function of the gate potential is $V_g^0=2e/C$. These oscillations arise from the semiclassical tunneling of the superconducting phase on the grain, and are therefore exponentially small in $\sqrt{E_C/\Delta}$ at $E_C \ll \Delta$. In addition, we predict a crossover from adiabatic to diabatic tunneling at the backscattering probability $r_{\rm ad} \sim \sqrt{E_C/\Delta}$. At backscattering below $r_{\rm ad}$, the amplitude ε of the Coulomb oscillations is proportional to the square root of the lesser (of the two contacts) reflection probability $\sqrt{r_{\rm min}}$. This is in contrast to the case of a normal double-contact system,[20] in which ε is proportional to the product $\sqrt{r_1 r_2}$.

The average Josephson current–phase relation $I(\alpha)$ is shown to be strongly nonsinusoidal and roughly similar to the one known for a single nearly ballistic QPC in the sense that it exhibits abrupt ''switching'' between positive and negative values of the current as the phase varies via $\alpha=\pi$. The new feature of our system is that it is possible to vary the width of the switching region $\delta\alpha$ by the electric gate potential V_g; in the case of equal reflection probabilities $r_1=r_2$, this electric modulation is especially pronounced, $\delta\alpha \propto |\cos(\pi C V_g/2e)|$. The noise spectrum of the supercurrent is found to consist mainly of two peaks: the ''zero-frequency'' peak due to rare thermal excitations of the upper level of the system, and another one centered around the energy difference ω_α between the two levels. The widths of these peaks are determined by the inverse lifetime τ of the two states of our TLS, which is due to electron-phonon and electromagnetic couplings. Both sources of level decay are expected to be very weak in the system considered, but the corresponding quantitative analysis is postponed to future studies; we present here only results for the *frequency-integrated* (over those narrow intervals $\sim 1/\tau$) noise power.

The $S-S-S$ device with almost ballistic contacts is a new type of system that can be used to implement an artificial ''spin 1/2''—an elemental unit for quantum computation. In comparison with conventional Josephson systems with tunnel junctions, which were proposed for use in adiabatic quantum computation,[11] the advantage of our system is that it can operate at considerably higher critical Josephson currents. Moreover, the current–phase characteristics of such a system is almost universal, in the sense that they are determined mainly by the microscopic parameters of the SC materials, and only weakly by the specifics of contact fabrication.

We are grateful to K. A. Matveev, Yu. V. Nazarov, and especially G. B. Lesovik for many useful discussions. The research of M. V. F. was supported by INTAS-RFBR Grant No 95-0302, collaboration grant No. 7SUP J048531 from the Swiss National Science Foundation, DGA Grant No. 94-1189, and the Program ''Statistical Physics'' of the Russian Ministry of Science.

[1] L. I. Glazman and K. A. Matveev, Zh. Éksp. Teor. Fiz. **98**, 1834 (1990) [Sov. Phys. JETP **71**, 1031 (1990)].
[2] K. A. Matveev, Phys. Rev. B **51**, 1743 (1995).
[3] L. I. Glazman, K. A. Matveev *et al.*, Physica B **203**, 316 (1994).
[4] H. Takayanagi, T. Akazaki, and J. Nitta, Phys. Rev. Lett. **75**, 3533 (1995).
[5] C. W. Beenakker and H. van Houten, Phys. Rev. Lett. **66**, 3056 (1991).
[6] M. C. Koops *et al.*, Phys. Rev. Lett. **77**, 2542 (1996).
[7] A. Yu. Kasumov *et al.*, Phys. Rev. Lett. **77**, 3029 (1996).
[8] A. Yu. Kitaev, E-print quant-ph/9707021 (1997).
[9] D. P. Divincenzo, E-print cond-mat/9612126 (1996).
[10] A. Shnirman, G. Schoen, and Z. Hermon, E-print cond-mat/9706016 (1997).
[11] D. V. Averin, E-print quant-ph/9706026 (1997).
[12] A. Martin-Rodero, A. Levy Yeyati, and F. J. Garcia-Vidal, Phys. Rev. B **53**, R8891 (1996).
[13] D. Averin and H. T. Imam, Phys. Rev. Lett. **76**, 3814 (1996).
[14] G. B. Lesovik and A. Golubov, in *Proc. of the 31 Recontres de Moriond*, T. Martin *et al.* (eds.), Editions Frontieres, Gif-sur-Yvette (1996).
[15] C. W. Beenakker, Phys. Rev. Lett. **67**, 3836 (1991).
[16] W. Lichten, Atomic Physics, Vol. No. 4, G. zu Putlitz *et al.* (eds.), Plenum, New York (1975).
[17] A. I. Larkin and Yu. N. Ovchinnikov, Phys. Rev. B **28**, 6281 (1983).
[18] L. D. Landau and E. M. Lifshitz, Quantum Mechanics: Nonrelativistic Theory, Pergamon Press, Oxford (1977).
[19] A. Martin-Rodero *et al.*, E-print cond-mat/9607079 (1996).
[20] A. Furusaki and K. A. Matveev, Phys. Rev. B **52**, 16676 (1995).
[21] G. B. Lesovik and R. Loosen, JETP Lett. **65**, 295 (1997); G. B. Lesovik, in *Proceedings of the Conference ''Meso-97''* Chernogolovka; submitted to Usp. Fiz. Nauk (1998).

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

Effect of isotopic composition on the linear thermal expansion coefficient of a germanium crystal lattice

A. P. Zhernov

Russian Research Center ''Kurchatov Institute,'' 123182 Moscow, Russia
(Submitted 25 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 654–668 (August 1998)

Features of the thermal expansion coefficient $\alpha(T)$ of crystal lattices with different isotopic compositions have been analyzed. The case of germanium lattices has been studied in detail.
 [S1063-7761(98)02108-8]

1. INTRODUCTION

Many problems concerning thermal expansion of crystal lattices have been studied in detail (see, for example, the monograph by Novikova[1] and review by Barron *et al.*[2]). Nonetheless, the thermal expansion coefficient α of crystals with different isotopic compositions, to the best of the author's knowledge, has not been investigated. For this reason, the paper considers effects of the first and second order in the difference between isotope masses on the expansion coefficient α.

The study is based on the standard quasiharmonic approximation. In other words, the temperature dependence of the lattice constant a is calculated. The point is that in real materials the energy of interatomic interaction, related dynamic parameters, and frequencies ω_l of phonon modes (here the index l labels different modes) are functions of $a(T)$ because of the lattice anharmonicity, whereas in the harmonic approximation the frequencies ω_l are independent of $a(T)$. In the quasiharmonic approach used in this study, the terms of order higher than two in the potential energy expansion in powers of displacement coordinates are rejected. In this case, some of anharmonicity effects are taken into account since a is a function of T and ω_l depends on a.

The paper considers a specific example of changes in α with variations of the isotopic composition of germanium crystals over a wide temperature range. The case of germanium is interesting for the following reason. It is known[1,2] that the temperature dependence of the thermal expansion coefficient α is determined in most cases by the temperature dependence of the lattice specific heat C_L. This is because the partial Grüneisen parameters γ_l, which characterize ω_l as a function of volume, are usually close to their average values in the high-temperature range. Since γ_l is an alternating function of l, the function $\alpha(T)$ in Ge is largely determined by the total Grüneisen parameter $\gamma(T)$, but not the lattice specific heat C_L. Therefore changes in the thermal expansion of Ge, unlike that of standard materials, should be, first of all, controlled by $\gamma(T)$. This paper analyzes the unconventional temperature dependence of α, taking as an example a comparison between calculations for highly enriched ^{70}Ge, ^{74}Ge, and ^{76}Ge crystals, and a crystal with the natural isotopic composition with the average atomic mass $M = 72.59$.

This research has been stimulated by investigations of chemically pure and isotopically enriched germanium crystals performed by the group directed by Ozhogin. They have measured the thermal conductivity of these materials,[3] and an investigation of other properties is under way.

2. THERMAL EXPANSION AND DENSITY OF PHONON STATES IN A CRYSTAL WITH ISOTOPIC DISORDER

2.1. Thermal expansion

It is known that the crystal free energy is expressed by

$$F = E_{\rm el} + F_{\rm v}, \tag{1}$$

where $E_{\rm el}$ is the equilibrium potential energy, and $F_{\rm v}$ is the contribution from vibrational energy. In the quasiharmonic approximation, the energy $F_{\rm v}$ is expressed in terms of the density of phonon states g:

$$F_{\rm v} = \int_0^\infty d\omega \omega f_T(\omega) g(\omega),$$

$$f_T(\omega) = \frac{-T}{\omega} \ln[n(\omega)+1] + \frac{1}{2}, \tag{2}$$

where $n(\omega) = (e^{\omega/T} - 1)^{-1}$. In order to simplify the equations, we set the Boltzmann and Planck constants equal to unity.

Using expressions for the free energy (1) and (2), one can derive the unit cell volume $\Omega = \Omega(T)$ at temperature T from the equilibrium condition. For a cubic crystal we have

$$\partial F(\Omega)/\partial\Omega = 0. \tag{3}$$

Suppose that we know the crystal parameters at a certain temperature $T_0 < T$. For simplicity we are considering the case of a cubic lattice (in this case thermal expansion is isotropic). With due account of these conditions, let us expand F about the equilibrium state, which is characterized by the volume $\Omega_0 = \Omega(T_0)$. In what follows, we take $T_0 = 0$. We have

$$F(\Omega)=F(\Omega_0)+\frac{1}{2}\left(\frac{\partial^2 E_{\rm el}}{\partial\Omega^2}\right)_{\Omega_0}(\Omega-\Omega_0)^2 + \left(\frac{\partial F_{\rm v}}{\partial\Omega}\right)_{\Omega_0}(\Omega-\Omega_0)\,. \tag{4}$$

From Eqs. (3) and (4) we directly derive a standard expression for the change in the cell volume $\Omega-\Omega_0$ as a function of temperature:

$$\frac{\Omega-\Omega_0}{\Omega_0}=-\frac{(\partial F_{\rm v}/\partial\Omega)_{\Omega_0}}{B_0},\quad B_0=\left(\Omega\frac{\partial^2 E_{\rm el}}{\partial\Omega^2}\right)_{\Omega_0}. \tag{5}$$

Here B_0 is the bulk modulus at $T=T_0$ (see also Refs. 4–6).

Next, the parameter $\Omega-\Omega_0$ is expressed directly in terms of the difference $\Delta a=a-a_0$, where a and a_0 are the lattice constants at T and T_0. We have

$$\frac{a-a_0}{a_0}=\frac{\Omega-\Omega_0}{3\Omega_0}+O((\Delta a)^2). \tag{5$'$}$$

The linear thermal expansion coefficient $\alpha(T)$ of a cubic lattice is

$$\alpha(T)=\frac{\partial}{\partial T}\frac{a-a_0}{a_0}. \tag{6}$$

The temperature dependence of α is determined by $\partial F_{\rm v}/\partial\Omega$. Note that the expression for the density of states g in the formula for $F_{\rm v}$ can be written as a power series in the parameter which is the difference between the isotope masses.

Note also that in classical statistical mechanics the partition function Z, which contains Gibbs's distribution function and depends on the atomic displacements and momenta, can be expressed as a product of the integrals J_1 and J_2. It is well known that the former is determined only by the distribution with repect to the components of atomic momenta (i.e., the kinetic energies of the atoms). The second integral is a function of the atomic coordinates and is determined by the potential energy of dynamic interatomic interaction through atomic force constants. The integral J_1, which depends on the kinetic energies of the isotopes and their masses, can be easily calculated and is independent of the cell volume. On the other hand, the integral J_2 is a function of cell volume, since the dynamic parameters depend on the lattice constants. Since, by definition,

$$F_{\rm v}=-T\ln Z=-T(\ln J_1+\ln J_2),$$

the derivative $(\partial F_{\rm v}/\partial\Omega)_{\Omega_0}$ does not depend on atomic mass. In other words, the thermal expansion coefficient is not sensitive to the isotope composition in the case of classical statistical mechanics. Changes in $\alpha(T)$ due to variations in the isotopic composition of a crystal are due to quantum effects.

2.2. Density of states in a crystal with isotopic disorder

Let us calculate the density of vibrational states in a crystal containing different isotopes. The Hamiltonian of such a crystal is expressed by

$$H=\frac{1}{2}\sum_{\mathbf{s}}\frac{p_{\mathbf{s}}^2}{2M_{\mathbf{s}}}+\frac{1}{2}\sum_{\mathbf{ss}'}\Phi_{\mathbf{ss}'}u_{\mathbf{s}}u_{\mathbf{s}'}\,. \tag{7}$$

Here $u_{\mathbf{s}}$ and $p_{\mathbf{s}}$ are the dynamic displacement and momentum operators of the atom located at site $\mathbf{s}$ of mass $M_{\mathbf{s}}$.

The expression for the vibrational energy of a system characterized by Eq. (7) can be written, with due account of the virial theorem, in the form

$$E_{\rm v}=\sum_{\mathbf{s}}\frac{\langle p_{\mathbf{s}}^2\rangle}{M_{\mathbf{s}}}=\sum_{\mathbf{s}}M_{\mathbf{s}}\langle\dot{u}_{\mathbf{s}}^2\rangle\,, \tag{8}$$

where angular brackets denote thermodynamic averaging. After averaging Eq. (8) over various isotopic configurations, we find

$$\overline{E}_{\rm v}=M_c\sum_{\mathbf{s}}\langle\langle\dot{u}_{\mathbf{s}}^2\rangle\rangle_c+\sum_{\mathbf{s}}(M_{\mathbf{s}}-M_c)\langle\langle\dot{u}_{\mathbf{s}}^2\rangle\rangle_c\,. \tag{9}$$

Here $\langle\ldots\rangle_c$ denotes configurational averaging, M_c is the effective average atomic mass, $M_c=\Sigma_{\mathbf{s}}M_{\mathbf{s}}/N$, and N is the total number of lattice sites.

Let us derive explicit expressions for the correlators in Eq. (9). To this end, we introduce a Green's function composed of operators of dynamic atomic diplacements:

$$G_{\mathbf{ss}'}(t)=-i\theta(t)\,\langle[u_{\mathbf{s}}(t),u_{\mathbf{s}'}(0)]\rangle\,. \tag{10}$$

Using Eq. (7), one can prove that the temporal Fourier transform of function (10) satisfies the equation

$$G_{\mathbf{ss}'}(\omega)=\overline{G}_{\mathbf{ss}'}(\omega)+\omega^2\times\sum_{\mathbf{s}_1}\overline{G}_{\mathbf{ss}_1}(\omega)(M_c-M_{\mathbf{s}_1})G_{\mathbf{s}_1\mathbf{s}'}(\omega)\,. \tag{11}$$

Here $\overline{G}(\omega)$ is the Green's function of the ordered lattice with atoms of mass M_c at all its sites. By definition,

$$\overline{G}_{\mathbf{ss}'}(\omega)=\frac{1}{M_cN}\sum_{\mathbf{f}j}\frac{\exp(i\mathbf{f}\cdot(\mathbf{r}_{\mathbf{s}}-\mathbf{r}_{\mathbf{s}'}))}{\omega^2-\omega^2(\mathbf{f},j)}, \tag{12}$$

where $\mathbf{r}_{\mathbf{s}}$ is the radius-vector of site $\mathbf{s}$, $\omega(\mathbf{f},j)$ denotes the frequency of the phonon mode with quasimomentum $\mathbf{f}$ and polarization j. To simplify our formulas in what follows, we introduce the notation $l=(\mathbf{f},j)$. In addition, we will ignore the presence of polarization vectors in Eq. (12) for the Green's functions.

Using the iteration method, we express an approximate solution to Eq. (11) in the form[7]

$$G_{\mathbf{ss}'}(\omega)\approx\overline{G}_{\mathbf{ss}'}(\omega)+\omega^2\sum_{\mathbf{s}_1}\overline{G}_{\mathbf{ss}_1}(\omega)(M_c-M_{\mathbf{s}_1})\overline{G}_{\mathbf{s}_1\mathbf{s}'}(\omega) + \omega^4\sum_{\mathbf{s}_1\mathbf{s}_2}\overline{G}_{\mathbf{s}_1\mathbf{s}_2}(\omega)(M_c-M_{\mathbf{s}_1})\overline{G}_{\mathbf{s}_1\mathbf{s}_2}(\omega) \times(M_c-M_{\mathbf{s}_2})\overline{G}_{\mathbf{s}_2\mathbf{s}'}(\omega). \tag{13}$$

From Eq. (13) we derive the Green's function averaged over isotopic configurations, $\langle G_{\mathbf{ss}'}\rangle_c$. We have

$$\langle G_{\mathbf{ss}'}(\omega)\rangle_c \approx \overline{G}_{\mathbf{ss}'}(\omega) + \omega^4 M_c^2 \xi^2 \overline{G}_c(\omega) \sum_{\mathbf{s}_1} \overline{G}_{\mathbf{ss}_1}(\omega)\overline{G}_{\mathbf{s}_1\mathbf{s}'}(\omega)\,. \tag{14}$$

Here $\overline{G}_c = \overline{G}_{\mathbf{ss}}(\omega)$, and the parameter ξ^2 is defined by the expression

$$\xi^2 = \frac{\langle M_{\mathbf{s}}^2\rangle - M_c^2}{M_c^2}, \quad \langle M_{\mathbf{s}}^2\rangle = \frac{1}{N}\sum_{\mathbf{s}} M_{\mathbf{s}}^2\,.$$

Let us take into account that

$$\begin{aligned} M_c^2 \operatorname{Im}\overline{G}_c(\omega) &\sum_{\mathbf{ss}_1} \overline{G}_{\mathbf{ss}_1}(\omega)\overline{G}_{\mathbf{s}_1\mathbf{s}}(\omega) \\ &= \overline{G}_c(\omega)\frac{1}{N^2}\sum_{\mathbf{ss}_1,\,\mathbf{ff}_1,j} \exp\{i\mathbf{f}\cdot(\mathbf{s}-\mathbf{s}_1)\} \\ &\quad\times\exp\{i\mathbf{f}_1\cdot(\mathbf{s}_1-\mathbf{s})\}\overline{G}_{\mathbf{f}j}(\omega)\,\overline{G}_{\mathbf{f}_1 j}(\omega) \\ &= \overline{G}_c(\omega)\sum_l \overline{G}_l(\omega)^2 \\ &= -\overline{G}_c(\omega)\frac{d}{d\omega^2}\sum_l \overline{G}_l(\omega), \end{aligned} \tag{15}$$

where the Green's function of the phonon mode l is

$$\overline{G}_l(\omega) = [\omega^2 - \omega^2(l) + i\delta]^{-1}.$$

Using Eqs. (14) and (15), we transform the first term in the expression for the average energy $\overline{E}_{\mathrm{v}}$ in Eq. (9):

$$\begin{aligned} M_c\sum_{\mathbf{s}} \langle\langle \dot{u}_{\mathbf{s}}^2\rangle\rangle_c \to \int_0^{\omega_{\max}} d\omega\left[n(\omega)+\frac{1}{2}\right]\omega^2\{&g_c(\omega^2) \\ &- \omega^4\xi^2[\operatorname{Re}\overline{G}_c(\omega)\operatorname{Im}\overline{G}'_c(\omega) \\ &+ \operatorname{Im}\overline{G}_c(\omega)\operatorname{Re}\overline{G}'_c(\omega)]\}. \end{aligned} \tag{16}$$

Here

$$g_c(\omega^2) = \sum_l \delta(\omega^2 - \omega^2(l)) \tag{17}$$

is the density function of squared frequencies in the ordered lattice of atoms with mass M_c, where the density of states is $g(\omega) = 2\omega g_c(\omega^2)$. The factor g_c describes the effects that are linear in the difference between the isotope masses, in addition, $\operatorname{Re}\overline{G}_c$ and $\operatorname{Im}\overline{G}_c$ in the equations above denote the real and imaginary parts of the lattice Green's functions:

$$\begin{aligned} \operatorname{Re}\overline{G}_c(\omega) &= \operatorname{Re}\langle G_{\mathbf{ss}}(\omega)\rangle_c = \frac{1}{M_c N}\sum_l \frac{1}{\omega^2-\omega^2(l)} \\ &= \int_0^{\omega_{\max}^2} dz^2 \frac{g_c(z^2)}{\omega^2 - z^2}, \end{aligned}$$

$$\operatorname{Im}\overline{G}_c(\omega) = \operatorname{Im}\langle G_{\mathbf{ss}}(\omega)\rangle_c = \frac{\pi}{M_c} g_c(\omega^2).$$

The primes in Eq. (16) denote the differentiation $d/d\omega^2$.

Consider the second term in expression (9) for the vibrational energy of the crystal. Using the approximate equation (13) for the Green's function, we obtain

$$\begin{aligned} \left\langle \sum_{\mathbf{s}} (M_{\mathbf{s}} - M_c)\langle \dot{u}_{\mathbf{s}}^2\rangle \right\rangle_c &\to \int_0^{\omega_{\max}} d\omega\omega^2\left[n(\omega)+\frac{1}{2}\right]\xi^2\omega^2 \\ &\times \operatorname{Im}\sum_{\mathbf{ss}_1,\mathbf{ff}_1,j} \exp\{i\mathbf{f}\cdot(\mathbf{s}-\mathbf{s}_1)\} \\ &\times\exp\{i\mathbf{f}_1\cdot(\mathbf{s}_1-\mathbf{s})\}\overline{G}_{\mathbf{f}j}(\omega)\,\overline{G}_{\mathbf{f}_1 j}(\omega) \\ &= -\int_0^{\omega_{\max}} d\omega\omega^2\left[n(\omega)+\frac{1}{2}\right]\xi^2\omega^2\frac{d}{d\omega^2}\operatorname{Im}\overline{G}_c(\omega). \end{aligned} \tag{18}$$

As a result, we obtain with due account of Eqs. (16) and (18) an expression for the energy $\overline{E}_{\mathrm{v}}$ of a crystal containing different isotopes in the form

$$\begin{aligned} \overline{E}_{\mathrm{v}} &= \int_0^{\infty} d\omega\omega^2\left[n(\omega)+\frac{1}{2}\right]\overline{g}(\omega^2), \\ \overline{g}(\omega^2) &= g_c(\omega^2) + \Delta g_c(\omega^2), \end{aligned} \tag{19}$$

where $\overline{g}$ is the desired density of states.

Recall that the factor g_c was defined above in Eq. (17), whereas the correction Δg_c to $\overline{g_c}$, which is quadratic in the difference between isotope masses due to disorder in their distribution, is given by the formula

$$\begin{aligned} \Delta g_c(\omega^2) = -\xi^2\omega^2\{\omega^2[&\operatorname{Re}\overline{G}_c(\omega)\operatorname{Im}\overline{G}_c(\omega) \\ &+ \operatorname{Re}\overline{G}'_c(\omega)\operatorname{Im}\overline{G}_c(\omega) + \operatorname{Im}\overline{G}'_c(\omega). \end{aligned} \tag{20}$$

Using Eqs. (17), (19), and (20) for the density of phonon states, we express the vibrational part F_{v} of the free energy defined by Eq. (2) to terms of first and second order in the difference between the isotope masses.

3. THERMAL EXPANSION, GRÜNEISEN PARAMETER, AND UNIT CELL VOLUME TO FIRST ORDER IN THE ISOTOPE MASS DIFFERENCE: BASIC RELATIONS

In the case of natural or enriched mixtures of isotopes of one element, such as germanium, the contribution to $\overline{g}$ of terms of the second order in the mass difference is essentially smaller than that of first-order terms. In what follows, we will neglect terms of the second order (they should be very important, for example, in ^{3}He and ^{4}He quantum crystals).

It is interesting that in the transport problem the situation is quite different. It is known that the phonon scattering time due to the isotopic disorder is inversely proportional to the isotope mass difference squared. The value of this parameter determines the amplitude of the thermal conductivity peak. Effects linear in the isotope mass difference also take place owing to the dependence of the Debye frequency ω_{D} on the average atomic mass, but their contribution is less important. The relationship between ω_{D} and M_c affects the freeze-out temperature of umklapp processes, and hence the location of the peak on the temperature axis.

Using the quasiharmonic approach and linearizing in the isotope mass difference, with due account of Eqs. (2) and

(17), we obtain for $F_{\rm v}$, which is the contribution of atomic vibrations to the free energy, a standard formula

$$F_{\rm v}=\frac{1}{2}\sum_l\ \omega(l)+T\sum_l\ \ln\left[1-\exp\left(-\frac{\omega(l)}{T}\right)\right]. \qquad (21)$$

The phonon frequencies in this equation are functions of the cell volume. Irrespective of the isotope composition, the following relation holds for each phonon mode l:

$$M_c\,\omega^2(l)=\phi(l), \qquad (22)$$

where $\phi(l)$ does not depend on M_c.

Note also that, in the case of a cubic crystal, the sum over wave vectors in the first Brillouin zone can be expressed in the form

$$\sum_{\mathbf{f},j}\ \ldots=\frac{\Omega}{2\pi^3}\sum_j\ \int d\mathbf{f}\ldots, \qquad (23)$$

where integration is performed over the first Brillouin zone, and Ω is the unit cell volume. By definition,

$$\Omega=Da^3,\quad \mathbf{f}=\pi\mathbf{Q}/a, \qquad (24)$$

where a is the lattice constant. The factor D is determined by the lattice configuration (bcc or fcc). It is important that $\mathbf{Q}$ is a dimensionless parameter (it depends on neither the cell volume nor the temperature).

It directly follows from Eqs. (23) and (24) that the vibrational component (21) of the free energy depends on the cell volume only through the frequencies of the phonon modes and is independent of the dimensions of the Brillouin zone (they are functions of temperature), i.e., integration limits. In other words, when Eq. (21) is applied to cubic crystals, it is enough to differentiate the integrand alone with respect to the volume. As a result, we obtain

$$\left(\frac{\partial F_{\rm v}}{\partial\Omega}\right)_{\Omega_0}=-\frac{1}{\Omega_0}\sum_l\ \gamma(l)\epsilon(\omega(l)),$$

$$\gamma(l)=-\frac{\partial\omega(l)/\partial\Omega}{\omega(l)/\Omega},\quad \epsilon(\omega(l))=\omega(l)\left(n(\omega(l))+\frac{1}{2}\right). \qquad (25)$$

Here $\gamma(l)$ is the partial Grüneisen factor for the lth phonon mode. It is introduced to take into account the different volume dependence in the frequencies of different modes. The function $\epsilon(l)$ denotes the contribution of a single mode to the thermal energy.

Now let us consider the thermal expansion coefficient $\alpha(T)$ of a cubic crystal lattice. Using Eqs. (5), (6), and (25), we obtain

$$\alpha(T)=-\frac{1}{3\Omega_0B_0}\frac{\partial}{\partial T}\sum_l\ \gamma(l)\epsilon(\omega(l))$$

$$=\frac{1}{3\Omega_0B_0}\sum_l\ \gamma(l)C_l(T), \qquad (26)$$

where

$$C_l(T)=\frac{\partial\epsilon(l)}{\partial T}=\frac{1}{T^2}\omega^2(l)n(\omega(l))[n(\omega(l))+1]$$

is the contribution of the lth phonon mode to the specific heat (see also Ref. 1).

If at small and large frequencies the factor $\gamma(l)$ is close to a certain mean value $\gamma_0=\text{const}$, we have instead of Eq. (26) the following expression for the thermal expansion coefficient:

$$\alpha(T)\approx\frac{1}{3\Omega_0B_0}\gamma_0C_L(T), \qquad (27)$$

where C_L is the lattice specific heat.

As for the total Grüneisen parameter $\gamma(T)$, it is defined as

$$\gamma(T)=\sum_l\ \gamma(l)C_l(T)/\sum_l\ C_l(T)\ . \qquad (28)$$

Note that $\gamma(T)$ is a weighted function of contributions from different modes.

Another parameter is considered concurrently with $\gamma(T)$:

$$\gamma_j(T)=\sum_{\mathbf{f}}\ \gamma(\mathbf{f}_{ij})C_{\mathbf{f}}(T)/\sum_l\ C_l(T), \qquad (29)$$

which is the total Grüneisen parameter for modes with polarization j.

Note that, since the range of populated modes narrows as the temperature drops, the Grüneisen parameter defined by Eq. (28) is a function of temperature owing to the difference between $\gamma(l)$ at high and low phonon frequencies.

As was noted above, the temperature dependence of the thermal expansion coefficient is determined in most cases by the specific heat, i.e., by Eq. (27). This is because the Grüneisen parameter and isothermal bulk modulus depend on temperature weakly in most cases, but, generally speaking, the relationship between γ and T is quite essential.

Now let us briefly discuss the general case of an anharmonic crystal. Owing to the interaction among phonons, the measured frequency $\tilde{\omega}_l$ of the lth mode is the sum of the harmonic part and anharmonic correction. Let us use the expression

$$\tilde{\omega}_l(\Omega,T)=\omega_l(\Omega)+\Delta_l^{(3)}+\Delta_l^{(4)},\quad \Omega=\Omega_0+\Delta\Omega\,. \qquad (30)$$

Here $\Delta_l^{(3)}$ and $\Delta_l^{(4)}$ are the anharmonic contributions due to standard anharmonic processes of the third and fourth orders.[8,9]

In view of this, it seems reasonable to define the appropriate partial Grüneisen parameter $\gamma_l^{(a)}$ as follows:

$$\gamma_l^{(a)}=-\{\partial\ln\tilde{\omega}_l(\Omega,T)/\partial\ln\Omega\}_{\Omega=\Omega(T)}, \qquad (31)$$

where $\Omega(T)$ is the cell volume with due account of thermal expansion. The expression for the anharmonic correction to the Grüneisen parameter γ_l has the form

$$\Delta\gamma_l^{(a)}(T)\approx\Delta\Omega(T)\frac{\partial\gamma_l}{\partial\Omega_0}-\frac{\partial\{(\Delta_l^{(3)}+\Delta_l^{(4)})/\omega_l\}}{\partial\Omega_0}. \qquad (32)$$

Now let us consider the case of two crystals with average atomic masses M_{c_1} and $M_c=M_{c_1}+\Delta M$. The change in

the thermal expansion coefficient $\Delta\alpha_c(T)=\alpha(M_c)-\alpha(M_{c_1})$ in the case of $|\Delta M|\ll M_{c_1}$ can be expressed by

$$\Delta\alpha_c \approx \alpha(M_{c_1})\left(\frac{\Delta\gamma_c}{\gamma(M_{c_1})}+\frac{\Delta C_c}{C_L}\right), \tag{33}$$

where

$$\Delta\gamma_c(T)=\gamma(M_c)-\gamma(M_{c_1}),$$

$$\Delta C_c(T)=C_L(M_c)-C_L(M_{c_1}).$$

Let us comment on Eq. (33). The shape of the temperature dependence of $\Delta\alpha_c$ is determined by the perturbation in the phonon spectrum. If the partial Grüneisen parameters γ_l vary slightly about certain mean values, changes in the phonon spectrum have little effect on the total parameter $\gamma(T)$, and the first term on the right-hand side can be neglected. If the difference between parameters γ_l at high and low frequencies is considerable, the terms with both $\Delta\gamma_c$ and ΔC_c in Eq. (33) are essential.

To conclude this section, let us discuss the dependence of the equilibrium cell volume (and lattice constant) on the average atomic mass of the mixture of isotopes. Given Eqs. (5), (25), and (22), we obtain at $T=0$

$$\delta\Omega=\frac{\Omega-\Omega_0}{\Omega_0}=\frac{1}{2B_0\Omega_0}\sum_l \gamma(l)\omega(l)\sim\frac{1}{\sqrt{M_c}}. \tag{34}$$

Note that the relationship between $\delta\Omega(T=0)$ and $\sqrt{M_c}$ is due to quantum oscillations of atoms in the ground state. The presence of the ground-state energy is a characteristic feature of quantum-mechanical motion. This reflects the fact that a state of absolute rest for a particle is meaningless in quantum mechanics.

Let us consider the case of two crystals with average atomic masses M_{c_1} and $M_c=M_{c_1}+\Delta M$, where $|\Delta M|\ll M_{c_1}$. Then it follows from Eq. (34) that the difference between the respective equilibrium cell volumes is

$$\delta\Omega(M_{c_1})-\delta\Omega(M_{c_1}+\Delta M)=-\rho\,\Delta M/M_{c_1},$$

$$\rho=\left(\frac{1}{2B_0\Omega_0}\sum_l \gamma(l)\omega(l)\right)_{M=M_{c_1}}.$$

In germanium, for example, we have $\rho\approx 2\times10^{-3}$, and $\Delta M/M_{c_1}$ can be up to several hundredths. Thus, relative changes in the cell volume at $T=0$ owing to variations in the isotopic composition can be about 10^{-5}.

At arbitrary temperatures, one can show, using Eq. (4), that the difference between equilibrium cell volumes for crystals with masses M_{c_1} and $M_c=M_{c_1}+\Delta M$ is

$$\Delta\bar{\Omega}=-\Delta M\left(\frac{1}{B}\frac{\partial^2 F_{\rm v}}{\partial\Omega\,\partial M}\right)_{M=M_{c_1}}.$$

Hence one can derive, using Eq. (21), the temperature dependence of $\Delta\bar{\Omega}$.

4. UNIVERSAL RELATIONSHIP FOR ISOTOPIC DEPENDENCE OF THERMAL EXPANSION COEFFICIENT

Taking into account the relationship $\omega(l)\propto M^{-1/2}$, one can determine the isotopic dependence of the thermal expansion factor described by Eq. (26). Let us specify some isotopic composition labeled by a subscript c_0. Then, by virtue of Eq. (22), we have for an arbitrary isotopic composition labeled by a subscript c a universal relationship like

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

It follows that, in the range of very low temperatures,

$$\alpha_c(T)=\beta_{c_0}(M_c/M_{c_0})^{3/2}T^3, \tag{35}$$

where β_{c_0} is independent of the atomic mass.

The factor in front of T^3 in Eq. (35) is sensitive to the isotopic composition and is almost linear in the difference between isotopic masses. (See also Ref. 10, which demonstrated that at low temperatures heavy isotopic impurities act as centers of anomalous thermal dilatation.) In the range of higher temperatures $T\gg\omega_l$, as directly follows from Eq. (26), where $C_l\approx 3[1-\omega_l^2/(12T^2)]$, thermal expansion should be a very weak function of the isotopic composition.

5. ISOTOPIC DEPENDENCE OF THERMAL EXPANSION COEFFICIENT AND THE GRÜNEISEN PARAMETER FOR GERMANIUM

On the base of the results described in the previous section, we have studied the isotopic dependence of $\alpha(T)$ and $\gamma(T)$ in germanium. In the process of our investigation, we have performed some calculations based on microscopic theories.

In particular, phonon mode frequencies ω_l have been determined on the base of the Born–Kármán theory. The interatomic interaction parameters have been determined by fitting to neutron diffraction data.[11] Further, the partial parameters γ_l were analyzed theoretically[12] with the help of the bond-charge model. It is based on the assumption that an electron charge concentrated at the center of a chemical bond can be treated as a dynamic variable affecting the interatomic interaction. This model can be applied to germanium because it allows one to describe a large flat region of the transverse acoustic phonon branch. It turned out that the interaction between nearest bond charges is stronger than between an atom and a neighboring bond charge. It is significant that the theory is in reasonable agreement with measurements of γ_l for a set of phonon modes.

In our study, we have used results of Refs. 11 and 12. In calculations of $\alpha(T)$ by [Eq. (26)], $\gamma(T)$ [Eq. (28)], and $\gamma_i(T)$ [Eq. (29)], integration was performed along symmetrical axes using the Houston formula (see Appendix) since the partial Grüneisen parameters were defined in Ref. 12 only along symmetrical directions. We have used the following parameters in our calculations: the lattice constant $a_0=5.658$ Å and bulk modulus $B_0=0.772\times10^{12}$ dyn/cm^2.

Note the following circumstance. The partial Grüneisen factors γ_l, by definition, are related to anharmonic atomic force constants. Consider as an example a linear chain with pair interaction between atoms. Then the expression for the

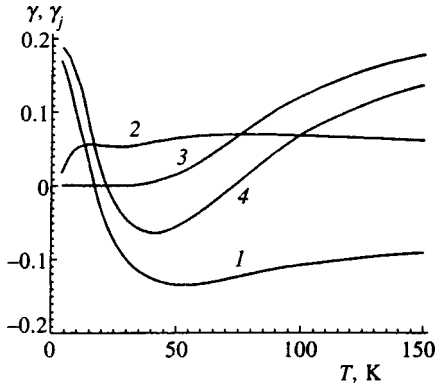


FIG. 1. Total Grüneisen parameters γ and γ_j as functions of temperature in the case of natural isotopic composition. Curves *1* and *2* plot partial total parameters for transverse and longitudinal acoustic modes, respectively. Curve *3* corresponds to optical modes. Curve *4* shows the total parameter γ.

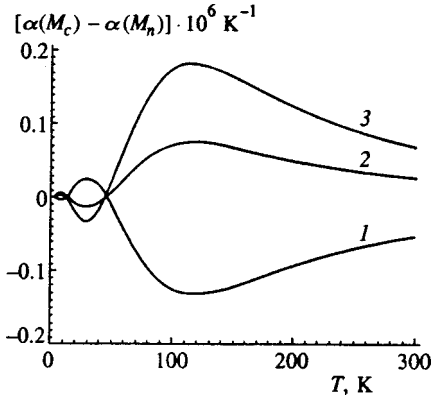


FIG. 3. Parameter $\alpha(M_c)-\alpha(M_n)$ versus temperature for the cases of light and heavy isotopes: (*1*) $M_c=70$; (*2*) $M_c=74$; (*3*) $M_c=76$.

interatomic forces f in the form of a series in powers of the atomic displacements from equilibrium positions has the form $f=-a_2u+a_3u^3+a_4u^4$. Here a_2, a_3, and a_4 are atomic force parameters of the second, third, and fourth orders. Consequently,

$$\gamma_l=-\frac{\partial \ln \omega_l}{\partial \ln \Omega}\propto-\frac{\partial f(u)/u}{\partial u}\approx a_3+a_4\,u.$$

This formula indicates that the Grüneisen parameter is really determined only by the anharmonic parameters a_3 and a_4. If the displacement amplitudes are small, we can take $\gamma_l \propto a_3$. In the general case, γ_l is a function of the displacement u, as a result, it depends not only on the volume, but also on the temperature. In our calculations we assumed that $\gamma_l \propto a_{3,l}$.

The results of our calculations are given in Figs. 1–4 and Table I.

Figure 1 shows curves of the Grüneisen parameter determined by Eq. (28) and partial Grüneisen parameters γ_j [Eq. (29)] for transverse (t) and longitudinal (ln) acoustic branches of the phonon spectrum, as well as for optical branches, versus temperature. For transverse acoustic and optical branches the factor γ_j is positive. In other words, the amplitudes of these phonons increase with interatomic distances and the frequencies drop. For acoustic transverse branches γ_t is negative. (This means that the attractive atomic forces, which determine parameters of transverse modes, increase faster than the repulsive forces when the interatomic distance changes.) The absolute value of γ_t is larger than that of γ_{ln}. Note also that the partial Grüneisen parameter for optical branches is ''unfrozen'' only at temperatures higher than that of liquid nitrogen, $T \geqslant T_{\mathrm{N}}$. In other words, the contribution of optical modes can be neglected at $T<T_{\mathrm{N}}$. As a result, the Grüneisen parameter γ summed over all branches of the phonon spectrum is negative at low temperatures and has a minimum at $T_m \approx 41$ K. At high temperatures $T>T_{\mathrm{N}}$ the total parameter γ is positive definite owing to the contribution of optical phonon modes.

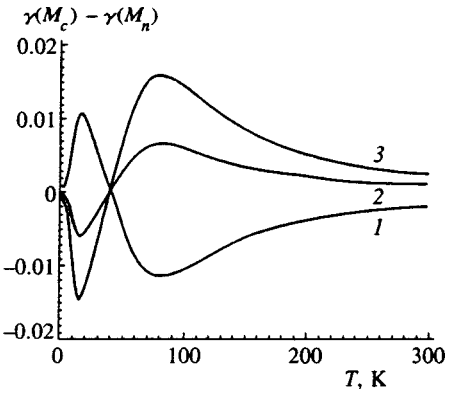


FIG. 2. Parameter $\gamma(M_c)-\gamma(M_n)$ at different M_c versus temperature: (*1*) $M_c=70$; (*2*) $M_c=74$; (*3*) $M_c=76$. In germanium $M_n=72.59$.

The comparison between calculations of $\gamma(T)$ and experimental data[1] indicates that the microscopic theory is in qualitative agreement with experiments. In the range of lower temperatures (on the left of $T_{\min}$), the agreement between the numbers is noticeably better.

Figures 2 and 3 illustrate the magnitudes of isotopic effects on $\alpha(T)$ and $\gamma(T)$ over a wide temperature range. They show plots of $\Delta\gamma_c=\gamma(M_c)-\gamma(M_n)$ and $\Delta\alpha_c=\alpha(M_c)-\alpha(M_n)$ versus temperature, where $M_n=72.59$ and $M_c=70$, 74, and 76. Recall that the natural isotopic composition corresponds to the mean atomic mass of 72.59. The masses of 70, 74, and 76 correspond to samples with maximum concentrations of specific isotopes, and samples of ^{70}Ge on one side and ^{74}Ge, ^{76}Ge on the other can be considered as materials with ''light'' and ''heavy'' isotopes in comparison with natural samples.

Now let us comment on the curves of $\Delta\gamma_c$ and $\Delta\alpha$ versus temperature. It is noteworthy that the curves are markedly nonmonotonic and have several extrema, in particular at $T_1 \approx 22$ K and $T_2 \approx 70$ K, where the calculated γ changes its sign. The values corresponding to light and heavy isotopes have opposite signs. In the range of high temperatures of about 300 K the isotopic effect is limited because the optical modes are not completely ''defrosted.'' Since the linear isotopic effect for the thermal expansion coefficient is determined by changes in both the Grüneisen parameter and specific heat, relative changes in $\Delta\alpha_c(T)$ are, generally speaking, larger than in $\Delta\gamma_c(T)$.

Figure 4 shows the curves of $\Delta\alpha_c/\alpha(M_n)$ versus temperature to illustrate the relative magnitudes of isotopic ef-

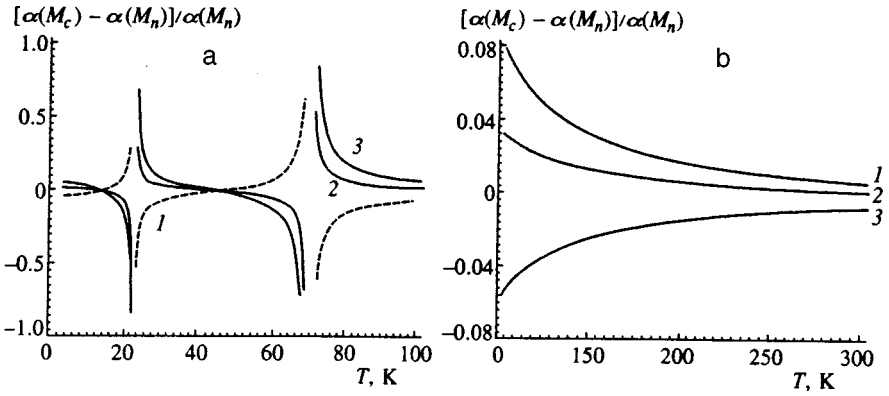


FIG. 4. Parameter $\alpha(M_c)/\alpha(M_n)-1$ as a function of temperature: (*1*) $M_c=70$; (*2*) $M_c=74$; (*3*) $M_c=76$.

fects for highly enriched crystals in comparison with natural samples. One can see in Fig. 4a that, in the temperature interval $T=5-100$ K, a strong isotopic effect (of about 10% or higher) occurs in narrow temperature ranges (≈ 5 K) near T_1 and T_2, where $\gamma(M_n)$ changes its sign. Figure 4b demonstrates that at relatively high temperatures the changes in $\alpha(T)$ in isotopically enriched samples can be several percent, and the absolute value of $\Delta\alpha_c/\alpha(M_n)$ drops with the temperature.

Finally, Table 1 lists values of $\Delta\gamma_c/\gamma(M_n)$ and $\Delta C_c/C_L$, which determine $\Delta\alpha_c/\alpha(M_n)$ [Eq. (26)], for $M_c=70$, 74 ($M_n=72.59$). As was noted above, the terms $\Delta\gamma_c/\gamma(M_n)$ due to the renormalization of the Grüneisen parameter have opposite alternating signs for light ($M_c=70$) and heavy ($M_c=74$) isotopes. On the contrary, the terms $\Delta C_c/C_L$ due to changes in the specific heat have the same sign throughout the studied temperature range, besides, the term $\Delta C_c/C_L$ is negative for the light isotope and positive for the heavy one. The values of $\Delta\gamma_c/\gamma(M_n)$ and $\Delta C_c/C_L$ are comparable far from points T_1 and T_2, where the Grüneisen parameter changes its sign. Around the points T_1 and T_2 the inequality $|\Delta\gamma_c/\gamma(M_n)| \gg |\Delta C_c/C_L|$ holds, i.e., the behavior of $\Delta\alpha_c/\alpha(M_n)$ is largely determined by the renormalization of the Grüneisen parameter.

TABLE I. Parameters $\Delta\gamma=\Delta\gamma(M_c)/\gamma(M_n)$ and $\Delta C=\Delta C(M_c)/C_L(M_n)$ as functions of T for crystals enriched in light (l, $M_c=$**70**) and heavy (h,$M_c=74$) isotopes.

T, K	$\Delta\gamma_l$	ΔC_l	$\Delta\gamma_h$	ΔC_h
2	0.0001	−0.0532	−0.0001	0.0294
10	0.0119	−0.0300	−0.0069	0.0169
15	0.0439	−0.0302	−0.0248	0.0161
20	0.1723	−0.0313	−0.0955	0.0137
30	−0.0434	−0.0197	0.0234	0.0119
40	−0.0033	−0.0166	0.0010	0.0094
42.5	0.0037	−0.0162	−0.0030	0.0090
45	0.0109	−0.0159	−0.0071	0.0088
47.5	0.0189	−0.0157	−0.0116	0.0086
50	0.0280	−0.0157	−0.0168	0.0085
60	0.0979	−0.0166	−0.0565	0.0081
67.5	0.4897	−0.0227	−0.2788	0.0062
80	−0.1331	−0.0131	0.0746	0.0091
90	−0.0695	−0.0135	0.0387	0.0085
100	−0.0465	−0.0130	0.0258	0.0079
150	−0.0150	−0.0089	0.0082	0.0050
200	−0.0076	−0.0058	0.0042	0.0033
250	−0.0047	−0.0040	0.0025	0.0022
300	−0.0032	−0.0029	0.0017	0.0016

Note also that in the temperature ranges $T_1<T<T_m(\gamma)$ and $T_2<T\leqslant 300$ K the terms $\Delta\gamma_c/\gamma(M_n)$ and $\Delta C_c/C_L$ are not only comparable, but also have the same sign. In the intervals 2 K$\leqslant T<T_1$ and $T_m(\gamma)\leqslant T<T_2$ these parameters have opposite signs and partially compensate one another. In the interval 2 K$\leqslant T<T_1$ the parameter $\Delta C_c/C_L$ dominates, and if $T_m(\gamma)\leqslant T<T_2$, the term $\Delta\gamma_c/\gamma(M_n)$ is more important.

Finally, anharmonic partial Grüneisen parameters have been estimated using Eqs. (30)–(32) based on the model of Ref. 12. It turned out that for most phonon modes $\gamma_l^{(a)}$ increase with temperature. The renormalization due to the anharmonicity at $T\approx 300$ K is 10–15%. At the same time, an anomalous situation takes place in the case of a soft transverse mode, namely, the absolute values of $\gamma_l^{(a)}$ diminish owing to the anharmonicity, for example, by a factor of about one and a half at room temperature. The main positive contribution is due to the anharmonic terms proper, i.e., the second term on the right-hand side of Eq. (32).

The effect of anharmonicity of higher orders has been estimated by replacing the cell volume with the measured values. Such corrections, generally speaking, are not small, but do not radically change the results, and the parameter $\gamma_l^{(a)}$ for the transverse mode, nonetheless, drops rapidly with the temperature.

Thus, taking into account the anharmonicity should lead to better agreement between calculations and conventional experimental data on thermal expansion of Ge.[1]

The role of anharmonicity in thermal expansion of germanium deserves a detailed investigation based on the microscopic model of bond charges.

6. CONCLUSIONS

In the reported work, the thermal expansion coefficient for isotopically mixed crystals has been determined with due account of terms of the first and second orders in the difference between isotope masses. A universal formula for α has been derived. An approximate description of the effects of atomic mass variation on $\alpha(T)$ and $\gamma(T)$ for germanium in a wide temperature range has been given. These parameters for natural and enriched isotope mixtures have been compared.

Unlike conventional crystals, the factor $\Delta\alpha_c/\alpha(M_n)$ in Ge as a function of temperature is largely determined by $\Delta\gamma_c/\gamma(M_n)$, i.e., by the renormalization of the Grüneisen parameter. This factor dominates over temperature ranges within several degrees of T_1 and T_2, where $\gamma(T)$ changes its sign. The parameter $\Delta C_c/C_L$, which usually determines $\Delta\alpha/\alpha(M_n)$, dominates only at very low temperatures $T<T_1$. At relatively high temperatures $T>T_2$ the effects of Grüneisen parameter renormalization and changes in the specific heat are comparable and have the same sign.

I am indebted to D. A. Zhernov for help in the work. Helpful remarks by S. M. Stishov are acknowledged. The work was supported by V. I. Ozhogin.

APPENDIX

Let us calculate the integral

$$J=\int_0^{\pi} d\vartheta \sin\vartheta / \int_0^{2\pi} d\phi I(\vartheta,\phi).$$

In accordance with the conditions of the problem, the function $I(\vartheta,\phi)$ has a cubic symmetry. Suppose that the integrand is defined along symmetrical directions (100), (110), and (111). Let us denote them as A, B, and C. Then, following Houston, we obtain an approximate expression[13]

$$J=\frac{4\pi}{35}(10J_A+16J_B+9J_C).$$

In the case of germanium discussed in this paper, the coordinates of high-symmetry points are

$$\Gamma=\frac{2\pi}{a}(0,0,0), \quad X=\frac{2\pi}{a}(0,0,1),$$

$$L=\frac{2\pi}{a}\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}\right), \quad K=\frac{2\pi}{a}\left(\frac{3}{4},\frac{3}{4},0\right).$$

Here Γ denotes the Brillouin zone center, X is the center of the square face of the unit cell, L is the center of the hexagonal face, K denotes the mid-points of edges of the hexagonal face, and a is the length of the edge of the unit lattice cube.

[1] S. I. Novikova, *Thermal Expansion of Solids* [in Russian], Nauka, Moscow (1974).
[2] T. H. Barron, J. G. Collins, and G. K. White, Adv. Phys. **29**, 609 (1980).
[3] V. I. Ozhogin, A. V. Inyushkin, A. N. Toldenkov, G. É. Popov, Yu. Kholler, and K. Ito, JETP Lett. **63**, 490 (1996)].
[4] G. Leibfried and E. Ludwig, *Theory of Anharmonic Effects in Crystals*, New York (1961).
[5] A. I. Ansel'm, *Introduction to Semiconductor Theory*, Mir, Moscow; Prentice-Hall, Englewood Cliffs, N.J (1981); Russian orig. Nauka, Moscow (1978).
[6] G. Leibfried and N. Brauer, *Point Defects in Metals. Introduction to the Theory*, Springer, Heidelberg (1972).
[7] Yu. M. Kagan and A. P. Zhernov, Zh. Éksp. Teor. Fiz. **53**, 1744 (1967) [Sov. Phys. JETP **26**, 999 (1968)].
[8] M. A. Krivoglaz, *Theory of X-ray and Thermal-Neutron Scattering by Real Crystals*, Plenum Press, New York (1969); Russian orig. Nauka, Moscow (1967).
[9] J. Reissland, *The Physics of Phonons*, Benjamin/Cummings, London (1973).
[10] Ya. A. Iosilevskii, Fiz. Tverd. Tela **9**, 2661 (1969) [Sov. Phys. Solid State **9**, 2090 (1969)].
[11] A. D. Zdetsis and C. S. Wang, Phys. Rev. **19**, 2999 (1979).
[12] R. Eryigit and I. P. Herman, Phys. Rev. B **53**, 7775 (1996).
[13] A. Maradudin, E. Montroll, and J. G. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation*, 2nd ed., Academic, New York (1971).

Translation provided by the Russian Editorial office.

High-frequency dielectric constant of a two-dimensionally disordered model medium

M. V. Éntin*)

Institute of Semiconductor Physics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 31 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 669–675 (August 1998)

An exact solution is derived for the two-dimensional two-phase model of a disordered medium proposed by Morozovskiĭ and Snarskiĭ [Ukr. Fiz. Zh. (Russ. Ed.) **28**, 1230 (1983)] that arises because of hierarchical mixing of phases with different dielectric constants. The problem reduces to a nonlinear recurrence relation for the dielectric constants. It is found that the expressions for the dielectric constants at the nth stage of the iteration process can be expressed in terms of elementary functions. It is also found that in the absence of absorption the high-frequency dielectric constant of a composite material with different signs of the initial dielectric constants ε does not converge to a limit and is an oscillatory function of the frequency and n. Finally, for massive and thin-film samples, the local plasmon frequencies in the medium are established. © *1998 American Institute of Physics.* [S1063-7761(98)02208-2]

Recent research has shown[2–7] that in randomly inhomogeneous macroscopic media made up of nondissipative microscopic parts, the values of high-frequency effective electric parameters never becomes stationary. In this respect the high-frequency case differs dramatically from the static limit. In Ref. 4, the renormalization equation for the hierarchical chain of impedances was studied by methods of dynamic-chaos theory and it was found that the impedance of a large chain is a fractal function of frequency. The absence of self-averaging leads to giant fluctuations of the high-frequency electric fields in a randomly inhomogeneous medium[2] and to divergence of the moments of the square of the absolute value of the electric field.[6] As a result there is giant enhancement of the nonlinear effects related to these moments, such as Raman scattering[5] and the photovoltaic effect.[7]

The work described in Ref. 2 was based on the Dykhne model[8] of a two-dimensional statistically equivalent two-phase random insulator medium. By analogy with Ref. 4, one should expect that in the nondissipative case the dielectric constant of a large system does not tend to a definite limit, the effective dielectric constant. Unfortunately, there is no way in which this can be established if the reasoning is based on the Dykhne model, which uses a limiting process. However, there exists a regular model of an isotropic two-dimensional medium, the Morozovskiĭ–Snarskiĭ (MS) model,[1] which in the static case has the same effective dielectric constant as the Dykhne medium. This model is based on a hierarchical construction, which facilitates the exact solution of the problem.

The MS model is based on a method of constructing a medium by connecting the initial phases in series and parallel: one takes extremely thin layers of equal thickness with conductivities σ_1 and σ_2 and stacks them. The resulting medium has an anisotropic conductivity with principal values σ_1' and σ_2'. At the next step of the hierarchy the procedure is repeated: the resulting medium is cut along the axes 1 and 2 into thin layers of equal thickness, which are then stacked. Figure 1 depicts two steps in the iteration, with one medium replaced by empty spaces. What we get is a chain of conductivities $\sigma_{1,2}^n$. Repeating this procedure indefinitely leads to equal values σ_1^∞ and σ_2^∞ that obey the Dykhne relations:

$$\sigma_1^\infty=\sigma_2^\infty=\sigma_{\text{eff}}=\sqrt{\sigma_1\sigma_2}. \tag{1}$$

It is more convenient to speak in terms of dielectric constants $\varepsilon=1+4\pi i\sigma/\omega$ instead of conductivities. The MS transformation for the complex-valued dielectric constants ε_1^n and ε_2^n of the system has the form

$$\varepsilon_1^{n+1}=\frac{\varepsilon_1^n+\varepsilon_2^n}{2},\quad \varepsilon_2^{n+1}=\frac{2\varepsilon_1^n\varepsilon_2^n}{\varepsilon_1^n+\varepsilon_2^n}. \tag{2}$$

In the dc case, ε_1 and ε_2 are purely imaginary and the resulting chain converges to the Dykhne result. The same is true in the case of purely real and positive ε_1 and ε_2, corresponding to a static dielectric constant.

By analogy with Ref. 4, one should expect that the map (2) leads to the dynamic-chaos scenario for purely real ε_1 and ε_2 with opposite signs. By computer simulation, Luk'yanets, Morozovskiĭ, and Snarskiĭ[9] showed that for $\varepsilon_1,\varepsilon_2<0$ the sequence of values of (2) does not converge. The result proves to be sensitive to the initial value, and there is no stage at which the system becomes isotropic: $\varepsilon_1^n\neq\varepsilon_2^n$.

In this paper we will obtain an exact solution for a finite-size MS model and study its behavior for complex-valued ε. The solution makes it possible to determine the plasmon spectrum of a finite-size MS model.

If we introduce the substitutions

$$\varepsilon_1^n=z_n\sqrt{\varepsilon_1\varepsilon_2},\quad \varepsilon_2^n=\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1^n}, \tag{3}$$

$$h=\frac{\varepsilon_1}{\varepsilon_2}, \tag{4}$$

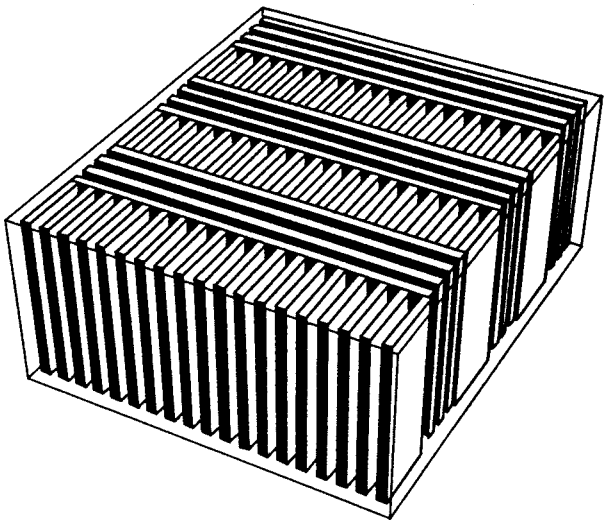


FIG. 1.

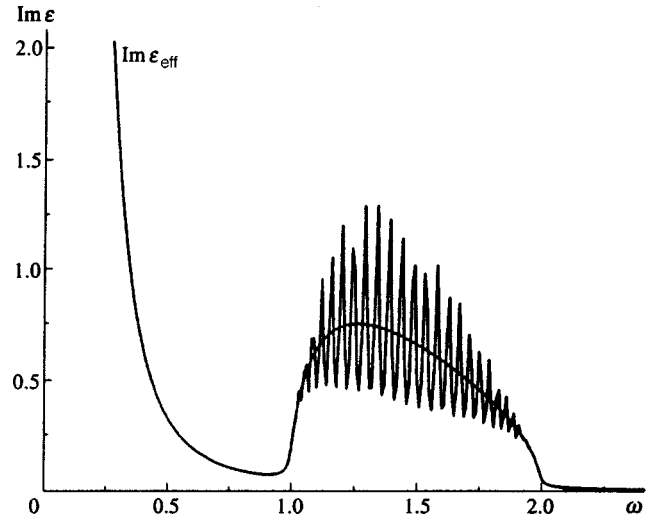


FIG. 2.

Eqs. (2) become

$$z_0=\frac{1}{\sqrt{h}},\quad z_{n+1}=\frac{z_n+1/z_n}{2}.$$

First we note that if we write $z_n=\tanh t$, the above recurrence relation becomes $z_{n+1}=\coth 2t$. Bearing this in mind, we obtain

$$z_n=\coth 2^n t,\quad n\geqslant 1. \tag{5}$$

The initial condition yields $t=\tanh^{-1} z_0=\coth^{-1}\sqrt{h}$. Let h be real and positive. In the limit $n\to\infty$,

$$z_n\sim 1+(-1)^n \exp(-2^n t).$$

This leads to the expression (1) for conductivity. The limit exists not only for h real but whenever t has a real part.

For h negative, by replacing hyperbolic functions by trigonometric functions we get

$$\varepsilon_1^n=-\sqrt{-\frac{\varepsilon_1}{\varepsilon_2}}\cot\left(2^n \cot^{-1}\sqrt{-\frac{\varepsilon_1}{\varepsilon_2}}\right),\quad \varepsilon_2^n=\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1^n}. \tag{6}$$

For n large and h negative, the z_n are rapidly oscillating functions of h. Within a small interval of values of h these functions behave as $\tan(2^n h+\text{const})$. For $h\sim 1$ the distance between the neighboring zeros or poles of z_n is of order $\pi 2^{-n}$, i.e., as n increases by one unit, the oscillation frequency doubles. Although such behavior is extraordinary, the function (5) is fairly regular, and no fractal behavior or alternating regular/fractal behavior, as in the dynamic-chaos pattern,[4] is observed. According to (3), the zeros of ε_1^n and the poles of ε_2^n (and vice versa) coincide with the zeros and poles of z_n, respectively.

The frequency dependence of the dielectric constant of a massive sample is determined in terms of h by the dependence on the frequency ω of the initial dielectric constants ε_1 and ε_2. Let us assume that the two initial media are metals described by the Drude–Lorentz model:

$$\varepsilon_1,\varepsilon_2=1-\frac{\omega_{p(1,2)}^2}{\omega(\omega+i/\tau_{1,2})}. \tag{7}$$

In the weak-absorption limit, $\tau\to\infty$ and

$$\varepsilon_1,\varepsilon_2=1-\frac{\omega_{p(1,2)}^2}{\omega^2}+i\,\frac{\omega_{p(1,2)}^2}{\omega^3\tau}.$$

For ε_1 and ε_2 to have different signs we must assume that the frequency lies in the interval between the plasma frequencies ω_{p1} and ω_{p2} of the massive materials; in particular, $\omega_{p1}<\omega<\omega_{p2}$.

Figure 2 depicts the frequency dependence of the imaginary part of the dielectric constant ε_1^6 at $2\omega_{p1}=\omega_{p2}=2$ and $1/\tau=0.02$, which on the average follows $\operatorname{Im}\varepsilon_{\text{eff}}=\operatorname{Im}\sqrt{\varepsilon_1\varepsilon_2}$ (the median line). Clearly visible in both curves are the low-frequency absorption region and the plasmon region $\omega_{p1}<\omega<\omega_{p2}$. In the latter region ε_1^6 oscillates. As $1/\tau$ decreases, the oscillations become more rapid and the sinusoidally modulated median curve splits into separate peaks. As n grows, the frequency of the peaks also increases, and the peaks merge into a median curve corresponding to the effective conductivity. Due to the merging of resonances, in the frequency range between the plasma frequencies the effective conductivity acquires (in a nondissipative medium) a finite real part.

According to Refs. 4 and 6, the resonances in the dielectric constant of a composite material reflect the local frequencies of oscillations in the medium. In a massive sample, the zeros of the dielectric constant determine the plasma frequencies of the sample in the long-wavelength limit. However, to determine the plasma frequencies we must know not only the microscopic properties of the medium but also the construction of the sample.

We start with a massive three-dimensional sample whose dielectric constant depends only on two coordinates, $(x,y)=\boldsymbol{\rho}$. The method by which the sample is constructed was described earlier. We assume that the hierarchical construction is terminated at the nth step; after that the medium is constructed homogeneously. The quantities ε_1^n and ε_2^n are

the diagonal elements of the tensor of the effective dielectric constant of the medium. Let the plasmon wave vector $\mathbf{k}$ lie in the plane $\boldsymbol{\rho}$ and be shorter than the reciprocal characteristic spatial scale of the penultimate hierarchy level. This is necessary if the medium is to be examined macroscopically.

With allowance for anisotropy, the Maxwell equations in the low-frequency limit,

$$\nabla\cdot\mathbf{D}=0, \quad \nabla\times\mathbf{E}=0, \tag{8}$$

yield a dispersion equation for plasmons with a wave vector $\mathbf{k}=(k_x,k_y)$ and polarization $\mathbf{E}\|\mathbf{k}$:

$$\varepsilon_1^n k_x^2+\varepsilon_2^n k_y^2=0.$$

Expressing the above formulas in terms of h and the angle α between $\mathbf{k}$ and the x axis, we find that

$$-h=\cot^2\ \gamma, \quad \omega^2=\omega_{p1}^2\ \sin^2\ \gamma+\omega_{p2}^2\ \cos^2\ \gamma,$$

where

$$\gamma=2^{-n}(\pi(j+1/2)\pm\alpha), \quad j \text{ is an integer.}$$

The number of different plasma modes is $2^{n+1}-1$, i.e., to within a number of order unity equals twice the number of elementary construction ''bricks.''

The result can easily be understood if we notice that at the boundary of two media with dielectric constants $\varepsilon_1,\varepsilon_2<0$ coupled plasmons with a frequency $\sqrt{(\omega_{p1}^2+\omega_{p2}^2)/2}$ emerge. The overlap of the fields generated by the different plasmons split their frequencies. At the same time, since the fields are assumed to be smooth in comparison to all the geometrical scales of the problem, all layers at a given partition level yield one additional frequency, and the number of resonant modes is independent of the number of layers into which the given level is partitioned.

Another example is presented by plasma oscillations in a two-dimensional composite film whose material is constructed in the same manner. Consider a film of thickness d that covers the surface $z=0$ of an insulator with a dielectric constant ε_0. We will assume that the film thickness exceeds the maximum size of the partition hierarchy but is smaller than the surface-plasmon wavelength. The first condition must be met if we want to replace the medium by an effective medium, and the second must be met if we want to solve the problem of the plasmon spectrum without taking into account the film thickness. The dielectric constant of the film is

$$\varepsilon_{1,2}^n(\mathbf{r})=d\varepsilon_{1,2}^n\delta(z).$$

The plasmon spectrum in the long-wavelength quasistatic approximation is given by Eq. (8) with allowance for the spatial dependence of the dielectric constant,

$$\varepsilon(\mathbf{r})=d\varepsilon_{1,2}^n\delta(z)+1+(\varepsilon_0-1)\theta(z),$$

where $\theta(z)$ is the Heaviside step function. Equation (8) for the surface plasmons in the film[10,11] has the following solution:

$$\phi=\phi_0 e^{i\mathbf{k}\cdot\boldsymbol{\rho}}e^{-k|z|}, \quad (\varepsilon_0+1)k=-4\pi d(\varepsilon_1^n k_x^2+\varepsilon_2^n k_y^2), \tag{9}$$

where ϕ is the plasmon frequency. In the limit $kd\ll 1$, Eq. (9) is satisfied as $\varepsilon_1^n k_x^2+\varepsilon_2^n k_y^2\to\infty$. Thus, the condition for the generation of plasma oscillations in a film is opposite to that for the generation of plasma oscillations in a massive sample. With the exception of angles α that are extremely small or very close to $\pi/2$, this condition implies that either $\varepsilon_1^n\to\infty$ or $\varepsilon_2^n\to\infty$. This yields the following dispersion relation:

$$\omega^2=\omega_{p1}^2\ \sin^2(2^{-(n+1)}\pi j)+\omega_{p2}^2\ \cos^2(2^{-(n+1)}\pi j).$$

Thus, the resonances in the dielectric constant determine the local plasmon frequencies. In the limit $n\to\infty$, the system under consideration is a unique example of an exactly solvable large non-one-dimensional system.

The results show that as $n\to\infty$, the neighboring resonances of the dielectric constant and, respectively, the plasmon frequencies are spaced almost evenly.

The local plasmon frequencies of a large system can be examined in the same way as electron states in a random medium. According to the Wigner–Dyson theory of random matrices,[12] in a system with nonseparable variables the spectrum of states is chaotic and the probability of neighboring states moving closer together behaves as a power law, $\propto\delta E^m$, where $m=1,2,4$, depending on the ensemble. The system we are considering here is two-dimensional and random, so that we can expect a similar behavior of the plasma resonances.

In one-dimensional quasiclassical wells the states in the spectrum are almost equally spaced, but the correlation in multidimensional systems is usually not so strict. For instance, the neighboring states of an electron in a rectangular two-dimensional well are not correlated. In an isotropic oscillator, which is an exception, the equal spacing of the spectrum arises because of degeneracy of the states, which is not the case in the present model.

In the hierarchical chain of impedances[4] described by a recurrence relation that cannot be made simpler, the impedance resonances are distributed at random with a power-like correlation. Computer simulation shows that the distribution function for the distances $\delta\omega$ between neighboring resonances behaves as $\delta\omega^{-1.25}$ over a broad range of values of $\delta\omega$. Probably, such power-law behavior is typical of systems with dynamic chaos.

However, due to the exact solvability of the map (5), no dynamic chaos emerges in the MS model, and the states make up an approximately equidistant spectrum.

I am grateful to A. A. Snarskiĭ for the allowing me to acquaint myself with his work prior to publication and to É. M. Baskin for fruitful discussions. The work was partially supported by grants from the Russian Fund for Fundamental Research (Grants Nos. 950204432, 960219353, and 97-02-18397) and Volkswagen-Stiftung.

*)E-mail: entin@isp.nsc.ru

[1] A. E. Morozovskiĭ and A. A. Snarskiĭ, Ukr. Fiz. Zh. [in Russian] **28**, 1230 (1983).
[2] F. Brouers, S. Blacher, and A. K. Sarychev, in *Fractal Reviews in the Natural and Applied Sciences*, M. M. Novak (Ed.), Chapman & Hall,

London (1995), p. 237; F. Brouers, S. Blacher, N. Henrioulle, and A. K. Sarychev, in *Electrical Transport and Optical Properties of Inhomogeneous Media*, A. M. Dykhne, A. N. Lagarkov, and A. K. Sarychev, Scientific Center for Applied Problems in Electrodynamics, Moscow (1996), p. 46; A. N. Lagarkov, K. N. Rosanov, A. K. Sarychev, and N. A. Simonov, J. Phys. A **241**, 199 (1996).
[3] J. P. Clerc, G. Giraud, J. M. Laugier, and J. M. Luck, Adv. Phys. **39**, 191 (1990).
[4] M. V. Éntin and G. M. Éntin, JETP Lett. **64**, 467 (1996).
[5] F. Brouers, S. Blacher, A. N. Lagarkov, A. K. Sarychev, P. Gadenne, and V. M. Shalaev, Phys. Rev. B **55**, 13 234 (1997).
[6] E. M. Baskin, M. V. Entin, A. K. Sarychev, and A. A. Snarskii, Physica A **242**, 49 (1997).
[7] M. V. Éntin, Fiz. Tekh. Poluprovodn. **31**, 973 (1997) [Semiconductors **31**, 829 (1997)].
[8] A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
[9] S. P. Luk'yanets, A. E. Morozovskiĭ, and A. A. Snarskiĭ (private communication).
[10] R. A. Ferrell, Phys. Rev. **111**, 1214 (1958).
[11] A. V. Chaplik, Zh. Éksp. Teor. Fiz. **60**, 1845 (1971) [Sov. Phys. JETP **33**, 997 (1971)].
[12] F. Dyson, J. Math. Phys. **3**, 140 (1962).

Translated by Eugene Yankovsky

High-frequency phenomena in organic conductors in a magnetic field

V. G. Peschanskiĭ*)

B. I. Verkin Low-Temperature Physicotechnical Institute, National Academy of Sciences, 310164 Kharkov, Ukraine

(Submitted 21 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 676–686 (August 1998)

The propagation of electromagnetic waves in organic layered conductors with metallic conductance and a quasi-two-dimensional electron energy spectrum of arbitrary form is studied theoretically. The depth of penetration of the electromagnetic field into the conductor is found, and it is discovered to be sensitive to the polarization of the incident wave. This is done for an arbitrarily oriented (with respect to the layers) magnetic field so strong that the radius of curvature of the electron path is much smaller than the electron's mean free path. It is established that studying these effects in experiments can reveal in detail the shape and dimensions of the Fermi surface and the relaxation properties of the conduction electrons.
 [S1063-7761(98)02308-7]

To a great extent the interest in low-dimensional organic conductors is due to their unusual behavior in strong magnetic fields and to a number of phase transitions at moderate pressures. The discovery of Shubnikov–de Haas oscillations of the magnetoresistance in tetracyano–tetracene halogens and a large family of tetrathiafulvalene-based ion-radical salts with charge transfer in magnetic fields of several rens of teslas suggests that these compounds have metallic conductivity and that the carrier mean free path l in them reaches several micrometers. At present strong magnetic fields can be generated inside such conductors, with the result that the curvature radius r of the conduction electron path becomes much smaller than l, so that the inverse problem of recreating the electron energy spectrum from the results of an experimental study of the kinetic phenomena in a magnetic field can be resolved.

Usually organic conductors are layered or filamentary structures with a marked electrical-conductivity anisotropy, and their electron energy spectrum is quasi-two-dimensional or quasi-one-dimensional. The topologically simplest model of a Fermi surface for quasi-two-dimensional conductors, a slightly fluted cylinder, is in good agreement with the results of experimental studies of the magnetoresistance and the Shubnikov–de Haas effect of the organic conductors $(BEDT\text{-}TTF)_2I_3$ and $(BEDT\text{-}TTF)_2IBr_2$ (Refs. 1–6). However, the unusual behavior of the magnetoresistance of the family of salts of the $(BEDT\text{-}TTF)_2MHg(SCN)_4$ type,[7–15] where M=(K, Rb, or Tl), suggests that the Fermi surface of such layered conductors is complicated, and to reproduce the electron energy spectrum of such compounds one must employ, in addition to galvanomagnetic results, alternative methods of studying the carrier dispersion law. A possible topological structure of the electron energy spectrum of this family of organic compounds, which follows from band calculations,[16,17] is a Fermi surface that, in addition to a slightly fluted cylinder, contains two quasi-one-dimensional sheets in the form of slightly rippled planes on which the charge carrier velocity has a preferred direction in the plane of the layers. The extent to which such a model spectrum corresponds to the real dispersion law for the charge carriers can easily be established by experimentally studying the high-frequency phenomena in a strong magnetic field ($r \ll l$), which will make it possible to determine the fine details of the Fermi surface and the relaxation properties of the charge carriers in such conductors.

In this connection we consider the propagation of electromagnetic waves in layered conductors whose Fermi surface consists of a quasi-two-dimensional cavity and a quasi-one-dimensional cavity, i.e., the electron energy spectrum consists of two bands with quasi-two-dimensional,

$$\varepsilon(\mathbf{p}) = \sum_{n=0}^{\infty} \varepsilon_n(p_x, p_y) \cos\left[\frac{anp_z}{\hbar} + \alpha_n(p_x, p_y)\right], \tag{1}$$

and quasi-one-dimensional,

$$\varepsilon'(\mathbf{p}) = \sum_{\substack{nm \\ q=0}}^{\infty} A_{nmq} \frac{\cos a_1 np_x}{\hbar} \frac{\cos a_2 mp_y}{\hbar} \frac{\cos aqp_z}{\hbar}, \tag{2}$$

dispersion laws for the charge carriers; the relationship between the numbers of conduction electrons in each energy band is taken to be arbitrary. Here a is the distance between the layers, a_1 and a_2 are the crystal lattice periods in the layer planes, and $\alpha_n(p_x, p_y) = -\alpha_n(-p_x, -p_y)$. In expression (1) for the quasi-two-dimensional carrier spectrum, the coefficients of the cosines rapidly decrease with increasing number n, so that the maximum value of the function $\max[\varepsilon_F - \varepsilon_0(p_x, p_y)] = \eta\varepsilon_F$ at the Fermi surface $\varepsilon(\mathbf{p}) = \varepsilon'(\mathbf{p}) = \varepsilon_F$ is much smaller than the Fermi energy ε_F, i.e., $\eta \ll 1$. At the same time we assume that η is not too small, so that $\eta \gg \hbar\Omega/\varepsilon_F$, where $\Omega = eH/m^*c$ is the gyration frequency of the electron in a magnetic field along a closed orbit, and m^* is the electron cyclotron effective mass. This condition makes it possible to use the semiclassical approximation to describe nonequilibrium processes in conductors with a quasi-two-dimensional electron energy spectrum. In the dispersion law $\varepsilon'(\mathbf{p})$ for charge carriers with a quasi-one-

dimensional energy spectrum, the terms with $m=q=0$ provide the main contribution. If we assume that $A_{100}=U\geqslant\varepsilon_F-A_{000}$ holds and all the other coefficients A_{nmq} vanish, the hollow of the Fermi surface corresponding to this energy band consists of two planes,

$$p_x=\pm\frac{\hbar}{a_1}\arccos\frac{\varepsilon_F-A_{000}}{U}.$$

The slight rippling of these planes can be taken into account by assuming that two more terms in (2) are finite, $A_{010}=\eta_1 U$ and $A_{001}=\eta_2 U$, where not only η_2 but also η_1 must be much smaller than unity.

The complete system of equations describing the propagation of electromagnetic waves in conducting media consists of the Maxwell equations

$$\operatorname{curl}\mathbf{H}=-i\omega\mathbf{E}+\frac{4\pi j}{c}, \tag{3}$$

$$\operatorname{curl}\mathbf{E}=i\omega\mathbf{B},\quad \mathbf{B}=\mathbf{H}+4\pi\mathbf{M} \tag{4}$$

and the kinetic equation for the charge-carrier distribution function,

$$f(\mathbf{p},\mathbf{r},t)=f_0(\varepsilon)+\psi(\mathbf{p},\mathbf{r})e^{-i\omega t}\frac{\partial f_0(\varepsilon)}{\partial\varepsilon}, \tag{5}$$

which makes it possible to relate the current density

$$\mathbf{j}(\mathbf{r})=-\int\frac{2d^3p}{(2\pi\hbar)^3}e\mathbf{v}\psi(\mathbf{p},\mathbf{r})\frac{\partial f_0(\varepsilon)}{\partial\varepsilon} \tag{6}$$

to the electric field $\mathbf{E}(\mathbf{r})$ of the wave.

In Eqs. (4), $\mathbf{M}$ is the magnetization of the conductor. The magnetic susceptibility $\chi_{ij}=\partial M_i/\partial B_j$ may reach values of order unity at low temperatures, where the de Haas–van Alphen effect manifests itself most clearly. Under these conditions the homogeneous state proves to be unstable[18,19] and small-amplitude nonlinear waves can be excited.[20] When the temperature of a conductor in which there is no magnetic ordering is not too low, the magnetic susceptibility is negligible and there is no need to distinguish between the magnetic field $\mathbf{H}$ and the magnetic induction $\mathbf{B}$. This approximation will be used here. We will also assume that the perturbation of the system of charge carriers by the electromagnetic wave is weak, so that we can work to first order in the weak electric field of the wave.

The Maxwell equations in this approximation are linear, and it is sufficient to keep only one temporal harmonic, i.e., the electromagnetic field can be assumed monochromatic with a frequency ω, a fact taken into account in the definition of the nonequilibrium addition to the Fermi distribution function $f_0(\varepsilon)$ of the charge carriers. Hence in the Maxwell equations (3) and (4) we have replaced the time derivatives of the electromagnetic fields by the factor $-i\omega$. In what follows, t is the time of the motion of a charge in a magnetic field according to the equation

$$\frac{d\mathbf{p}}{dt}=\frac{e}{c}\mathbf{v}\times\mathbf{H}. \tag{7}$$

An important assumption that makes it possible to represent the current density as (6) is the smallness of the quantum corrections to the electrical conductivity of the conductor, which oscillate with the magnetic field. If in the semiclassical approximation we ignore these corrections, the kinetic equation linearized with respect to the weak perturbation of the conduction electrons by the wave assumes the form

$$\left(\mathbf{v}\cdot\frac{\partial\psi}{\partial\mathbf{r}}-i\omega\psi+\frac{e}{c}(\mathbf{v}\times\mathbf{H})\cdot\frac{\partial\psi}{\partial\mathbf{p}}\right)\frac{\partial f_0(\varepsilon)}{\partial\varepsilon}+W_{\mathrm{col}}(\psi)$$
$$=e\mathbf{v}\cdot\mathbf{E}(\mathbf{r})\frac{\partial f_0(\varepsilon)}{\partial\varepsilon}, \tag{8}$$

where $W_{\mathrm{col}}(\psi)$ is the linear integral collision operator, which we take into account in the τ-approximation, i.e., we assume that $W_{\mathrm{col}}\{f\}=(f_0-f)/\tau$, with τ the carrier mean free time.

Under the conditions of the anomalous skin effect, when the depth δ to which an electromagnetic field penetrates the conductor (penetration depth) is smaller than the carrier mean free path, the manner in which the charge carriers are reflected by the sample surface $\mathbf{r}_s=0$ manifests itself most clearly, with the result that we must augment the kinetic equation with a boundary condition at the surface that scatters the conduction electrons:

$$\psi(\mathbf{p}_+,\mathbf{r}_s)=q(\mathbf{p}_-)\psi(\mathbf{p}_-,\mathbf{r}_s)+\int d^3pW(\mathbf{p},\mathbf{p}_+)$$
$$\times\{1-\theta(\mathbf{n}\cdot\mathbf{v})\}\psi(\mathbf{p},\mathbf{r}_s), \tag{9}$$

where the specular reflection parameter of the sample, $q(\mathbf{p})$, i.e., the probability that an electron incident on the surface $\mathbf{r}_s=0$ with a momentum $\mathbf{p}_-$ will be specularly reflected by the surface, is connected to the scattering indicatrix through the following relationship:

$$q(\mathbf{p}_-)=1-\int d^3pW(\mathbf{p},\mathbf{p}_+)\{1-\theta(\mathbf{n}\cdot\mathbf{v})\}, \tag{10}$$

where $\theta(\xi)$ is Heaviside's function, $\mathbf{n}$ is the normal to the surface of the conductor, and $\mathbf{p}_-$ and $\mathbf{p}_+$ (the momenta of the incident and reflected electrons, respectively) are connected by the specular reflection condition, which ensures that the energy and projection of momentum on the sample surface are conserved. When there are several groups of charge carriers, several specular reflection channels are possible, causing the enhancement within the bulk of the conductor of electromagnetic-field bursts[21] predicted by Azbel'.[22] The multichannel nature of the specular reflection of electrons from the sample boundary manifests itself most vividly in the transparency and surface impedance of a thin sample whose thickness L is less than or of order the carrier mean free path, but has no significant effect on the penetration depth for electromagnetic waves in a massive conductor ($L\gg l$). Hence we ignore band-to-band transitions of charge carriers when they scatter inside a massive sample and at the sample surface.

The integral term in the boundary condition (9) ensures that no current flows through the conductor's surface, but at high frequencies ω the solution of the kinetic equation is

weakly dependent on this functional,[23] so that if we ignore this term, in a magnetic field parallel to the sample surface when charge carriers with the quasi-two-dimensional spectrum (1) drift along open electron orbits into the interior of the conductor, the solution of Eq. (8) assumes the form

$$\psi(t_H,p_H,\mathbf{r})=\int_\lambda^{t_H} dt e\mathbf{v}(t,p_H)\cdot\mathbf{E}(\mathbf{r}(t,p_H)-\mathbf{r}(\lambda,p_H))\exp[\nu(t-t_H)]+q(\lambda,p_H)\times\{1-q(\lambda,p_H)\exp[\nu(2\lambda-T)]\}^{-1}\times\int_\lambda^{T-\lambda} dt e\mathbf{v}(t,p_H)\cdot\mathbf{E}(\mathbf{r}(t,p_H)-\mathbf{r}(\lambda,p_H))\exp[\nu(t-t_H+2\lambda-T)], \quad (11)$$

where $\nu=-i\omega+1/\tau$, $T=2\pi/\Omega$ is the period of motion of a charge along a closed orbit, $p_H=\mathbf{p}\cdot\mathbf{H}/H=\text{const}$, and λ is the root of the equation

$$\mathbf{r}(t,p_H)-\mathbf{r}(\lambda,p_H)=\int_\lambda^t \mathbf{v}(t',p_H)dt'=\mathbf{r}-\mathbf{r}_s \quad (12)$$

that is closest to t_H. Here and below the subscript ''H'' on a variable in momentum space that coincides with the time of motion of the charge in a magnetic field is dropped. For conduction electrons not colliding with the sample surface we must put $\lambda=-\infty$.

Let us now discuss the propagation of electromagnetic waves along the x axis in a magnetic field $\mathbf{H}=(0,H\sin\vartheta,H\cos\vartheta)$ parallel to the sample surface $x_s=0$. If the sample thickness is not only much larger than the mean free path of the charge carriers but also much larger than the penetration depth, then, to a high accuracy, the distribution of the variable electric field in the sample near the sample boundary has the same shape as in the half-space $x\geqslant 0$ occupied by the conductor. Following Reuter and Sondheimer,[24] we continue the electric field as an even function into the region of negative x and write the Fourier transforms for the current density and electric field:

$$\mathbf{j}(k)=2\int_0^\infty dx\,\mathbf{j}(x)\cos kx,$$

$$\mathbf{E}(k)=2\int_0^\infty dx\,\mathbf{E}(x)\cos kx. \quad (13)$$

Using the solution (11) of the kinetic equation, we can find the relationship between the Fourier transforms of the current density and those of the variable electric field:

$$j_i(k)=[\sigma_{ij}(k)+\sigma_{ij}^{(1)}(k)]E_j(k)+\int dk' Q_{ij}(k,k')E_j(k'), \quad (14)$$

where the contribution of the charge carriers with a quasi-two-dimensional energy spectrum to the high-frequency electrical conductivity is

$$\sigma_{ij}(k)=\frac{2e^3H}{c(2\pi\hbar)^3}\times\int dp_H\int_0^T dt\,v_i(t,p_H)\int_{-\infty}^t dt' v_j(t',p_H)\times\exp[\nu(t'-t)]\cos k[x(t',p_H)-x(t,p_H)]\equiv\langle e^2 v_i\hat{R}v_j\rangle, \quad (15)$$

and only conduction electrons obeying the dispersion law (2) contribute to $\sigma_{ij}^{(1)}(k)$.

The Maxwell equations in the Fourier representation,

$$\left(k^2-\frac{\omega^2}{c^2}\right)E_\alpha(k)-\frac{4\pi i\omega}{c^2}j_\alpha(k)=-2E'_\alpha(0),\quad \alpha=(y,z), \quad (16)$$

together with the material equation (14) make it possible to easily find the Fourier transforms of the variable electric field and then, using the inverse Fourier transform, the distribution of the electric field in the conductor.

The electric field $E_x(x)$ can be found by solving Poisson's equation

$$\text{div}\,\mathbf{E}=-4\pi\int\frac{2d^3p}{(2\pi\hbar)^3}e\psi(\mathbf{p},\mathbf{r})\frac{\partial f_0(\varepsilon)}{\partial\varepsilon}\equiv 4\pi e\langle\psi\rangle, \quad (17)$$

which in conductors with a high charge-carrier density asymptotically reduces to the electrical neutrality condition for such conductors:

$$\langle\psi\rangle=0. \quad (18)$$

The kernel $Q_{ij}(k,k')$ of the integral operator strongly depends on the state of the sample surface, i.e., on the probability of conduction electrons being specularly reflected by the surface. When the the charge carriers are reflected by the sample surface nearly specularly and the skin effect is anomalous, the second term on the right-hand side of Eq. (14) is the principal one for electrons with a quasi-two-dimensional energy spectrum, but is unimportant for electrons with a quasi-one-dimensional spectrum. A significant number of the charge carriers with a quasi-one-dimensional spectrum do not return to the sample surface; instead they are reflected from the surface into the bulk of the sample. The contribution of such carriers to the high-frequency current is only weakly sensitive to the state of the conductor's surface, i.e., to the form of the kernel $Q_{ij}(k,k')$. Allowance for the surface scattering of these electrons only modifies the numerical factor of order unity in the expression for the penetration depth for the electromagnetic wave and in no way influences the dependence of δ on the magnetic field and the low-dimensionality parameters of the energy spectrum of the conduction electrons. Hence, when calculating $\sigma_{ij}^{(1)}(k)$, we assume that the charge carriers with a quasi-one-dimensional spectrum of the form (2) are specularly reflected by the sample boundary.

When the magnetic field is so strong that the diameter $2r$ of the orbit of conduction electrons with a quasi-two-dimensional spectrum (1) is much smaller than the penetration depth, the contribution to the high-frequency current of

the charge carriers colliding with the conductor surface is negligible compared to the contribution of the electrons that do not reach the sample boundary. In this case we can limit ourselves to an approximation in which only the local coupling between the Fourier transforms of the current density and electric field is taken into account and the last term on the right-hand side of Eq. (14) is ignored. The penetration depth is easily found by solving the dispersion equation

$$\det[\,\delta_{\alpha\beta}-\xi\tilde{\sigma}_{\alpha\beta}(k)]=0,\quad \alpha,\beta=(y,z), \tag{19}$$

where $\xi=4\pi i\omega/(k^2c^2-\omega^2)$, and

$$\tilde{\sigma}_{\alpha\beta}(k)=\sigma_{\alpha\beta}(k)+\sigma^{(1)}_{\alpha\beta}(k) - \frac{[\sigma_{\alpha x}(k)+\sigma^{(1)}_{\alpha x}(k)][\sigma_{x\beta}(k)+\sigma^{(1)}_{x\beta}(k)]}{\sigma_{xx}(k)+\sigma^{(1)}_{xx}(k)}. \tag{20}$$

The contribution to $\tilde{\sigma}_{\alpha\beta}(k)$ of the charge carriers with a quasi-one-dimensional energy spectrum is primarily determined by the component $\sigma^{(1)}_{xx}(k)$, which to within small correction terms proportional to η_1^2 and η_2^2 has the form

$$\sigma^{(1)}_{xx}(k)=\sigma_1(k)=\frac{\sigma_1}{1+(kl_1)^2}, \tag{21}$$

where we have written $l_1=v_0\tau_1/(1-i\omega\tau_1)$, σ_1 is the contribution of this group of charge carriers to the electrical conductivity along the x axis in a uniform electric field, τ_1 is the mean free time of the charge carriers with the energy spectrum (2), and $v_0=(Ua_1/\hbar)\times\sin[(\varepsilon_F-A_{000})/U]$.

The magnetic-field dependence of $\sigma^{(1)}_{ij}(k)$ manifests itself only in the higher-order terms in the expansion in powers of the small parameters η_1 and η_2:

$$\sigma^{(1)}_{yy}(k)=\sum_{\pm}\frac{\eta_1^2\sigma_1 a_2^2U^2/4\hbar^2v_0^2}{1+(k\pm eHa_2\cos\vartheta/c\hbar)^2l_1^2}, \tag{22}$$

$$\sigma^{(1)}_{zz}(k)=\sum_{\pm}\frac{\eta_2^2\sigma_1 a^2U^2/4\hbar^2v_0^2}{1+(k\pm eHa\sin\vartheta/c\hbar)^2l_1^2}, \tag{23}$$

and allowing for these terms has no significant effect on the penetration depth for the electromagnetic fields.

The asymptotic behavior of the components $\tilde{\sigma}_{\alpha\beta}(k)$ in strong magnetic fields ($\gamma=1/\Omega\tau\ll1$),

$$\tilde{\sigma}_{yy}(k)=\frac{\sigma_1(k)(\gamma^2\sigma_0+\sigma_{zz}\tan^2\vartheta)+\gamma^2\sigma_0^2}{\sigma_1(k)+\gamma^2\sigma_0}, \tag{24}$$

$$\tilde{\sigma}_{yz}(k)=\tilde{\sigma}_{zy}(k)=\frac{\sigma_1(k)}{\sigma_1(k)+\gamma^2\sigma_0}\,\sigma_{zz}\tan\vartheta, \tag{25}$$

$$\tilde{\sigma}_{zz}(k)=\sigma_{zz}+\sigma^{(1)}_{zz}(k), \tag{26}$$

is extremely sensitive to the occurrence of a group of charge carriers with a quasi-one-dimensional energy spectrum. In Eqs. (24)–(26) we have left out the unimportant numerical factors of order unity and the small corrections of order $(kr)^2$ in the expression for σ_{zz}, and in σ_0 we discarded the contribution to the electrical conductivity parallel to the layers of charge carriers with the spectrum (1) at $H=0$.

If σ_1 and σ_0 are of the same order of magnitude, then $\tilde{\sigma}_{yy}(k)$ is much smaller than σ_0 over a broad range of magnetic field strengths, which makes the conductor much more transparent.

The dispersion equation (19) together with (24)–(26) make it possible to find the penetration depths for electromagnetic fields in a strong magnetic field:

$$\delta_1\simeq\frac{\delta_0}{\eta},\quad \delta_2\simeq\frac{\delta_0}{\gamma}, \tag{27}$$

where $\delta_0=\sqrt{c^2/2\pi\omega(\sigma_0+\sigma_1)}$.

In the limit $\sigma_1\ll\sigma_0$ but with $\sigma_1\geqslant\gamma^2\sigma_0$, the above expression for δ_2 acquires a small factor $\sqrt{\sigma_1/\sigma_0}$. For $\sigma_1\ll\gamma^2\sigma_0$ the penetration depths for the electric fields $E_z(x)$ and $E_y(x)$ differ considerably, and we have $\delta_z=\delta_1$ and $\delta_y\simeq\delta_0$, respectively; but if $\sigma_1\geqslant\gamma^2\sigma_0$ holds, the electric fields along the layers and perpendicular to the layers contain both components with very different penetration depths δ_1 and δ_2. Thus, even in fairly pure conductors, in the limit $l\eta\gg\delta_0$, not only $E_z(x)$ but also the electric field $E_y(x)$ decay in magnetic fields in which $r\ll\delta_0$ holds, over distances much larger than the carrier mean free path.

In the above formulas (24)–(27), $\tan\vartheta$ is not assumed to be very large, so that we have $\gamma=\gamma_0/\cos\vartheta\ll1$, where $\gamma_0=r_0/l$, and $r_0=c\hbar/eHa$ coincides in order of magnitude with the orbit radius of electrons with the spectrum (1) at $\vartheta=0$. We also assume that for $\eta\ll1$, $kr\ll1$, $\omega\tau\ll1$, and $\gamma\ll1$ the asymptotic expression for the electrical conductivity across the layers,

$$\sigma_{zz}(\eta,H)=\frac{2e^3H}{c(2\pi\hbar)^3}\int_0^{(2\pi\hbar/a)\cos\vartheta}\frac{dp_H(a/\hbar)^2}{1-\exp(-T/\tau)} \times\int_0^T dt\int_{t-T}^t dt'\sum_{n,m}\varepsilon_n(t,p_H)\varepsilon_m(t',p_H) \times\sin\left\{\frac{an}{\hbar}\left(\frac{p_H}{\cos\vartheta}-p_y(t,p_H)\tan\vartheta\right)\right\}nm \times\sin\left\{\frac{am}{\hbar}\left(\frac{p_H}{\cos\vartheta}-p_y(t',p_H)\tan\vartheta\right)\right\} \times\exp\left(\frac{t'-t}{\tau}\right) \tag{28}$$

has the form $\sigma_{zz}\simeq\eta^2\sigma_0$.

When the magnetic field is directed relative to the layers so that $\cos\vartheta\gg\gamma_0$ holds, the equations (7) for the charge motion imply that the period-average of the velocity of electrons with the quasi-two-dimensional spectrum (1), directed at right angles to the layers, is

$$\overline{v_z}(p_H,\vartheta)=\frac{1}{T}\int_0^T dt\,v_z(t,p_H,\vartheta), \tag{29}$$

and $p_x(t,p_H)$ and $p_y(t,p_H)$ are weakly dependent on the momentum projection $p_H=p_y\sin\vartheta+p_z\cos\vartheta$.

When ϑ is much larger than zero, there is always a $\vartheta=\vartheta_c$ (whose value is not unique) such that not only

$$\overline{v_z}(0,\vartheta)=\sum_{n=1}^{\infty}\frac{an}{\hbar T}\int_0^T dt\,\varepsilon_n(t,0)\sin\frac{anp_y(t,0)\tan\vartheta}{\hbar}=0, \tag{30}$$

but also

$$\sum_{n=1}^{\infty}\frac{an}{\hbar T}\int_0^T dt\,\varepsilon_n(t,0)\cos\frac{anp_y(t,0)\tan\vartheta}{\hbar}=0. \tag{31}$$

Clearly, at $\vartheta=\vartheta_c$ the value of $\overline{v_z}(p_H,\vartheta_c)$ decreases considerably, which leads to a significant change in the asymptotic behavior of the high-frequency conductivity component

$$\sigma_{zz}=\frac{e^2\tau am^*(\vartheta)\cos\vartheta}{8\pi^3\hbar^4}\sum_{n=1}^{\infty}n^2|I_n(\vartheta)|^2+\sigma_0\eta^2 \times\left[\eta^2 f_1(\vartheta)+\left(\frac{\gamma_0}{\cos\vartheta}\right)^2 f_2(\vartheta)+(kr)^2 f_3(\vartheta)\right], \tag{32}$$

where

$$I_n(\vartheta)=\frac{1}{T}\int_0^T dt\,\varepsilon_n(t)\exp\left[\frac{ianp_y(t)}{\hbar}\tan\vartheta\right], \tag{33}$$

and $f_i(\vartheta)$ is a function of order unity that depends on the specific form of the dispersion law for the charge carriers.

If $I_n(\vartheta)$ rapidly decreases with increasing n, the significant change in the asymptotic behavior of σ_{zz} for $\gamma_0\ll\cos\vartheta$ changes significantly when $I_1(\vartheta)$ vanishes, i.e., at values of ϑ at which $I_1(\vartheta_c)=0$. For such orientations of the magnetic field, not only does δ_2 increase with H but so does δ_1, reaching its limit δ_0/η^2.

For $\gamma_0\ll\cos\vartheta\ll 1$, the zeros of the function

$$I_1(\vartheta)=2\varepsilon_1(0)\sqrt{\frac{2\pi\hbar c}{aeHv_x'(0)\sin\vartheta}} \times\cos\left(\frac{aD_p\tan\vartheta}{2\hbar}-\frac{\pi}{4}\right) \tag{34}$$

recur with a period

$$\Delta(\tan\vartheta)=\frac{2\pi\hbar}{aD_p}, \tag{35}$$

where D_p is the diameter of the Fermi surface along the p_y axis, a prime indicates a derivative with respect to time t, and the origin of t is placed at the point of stationary phase, where $v_x=0$.

In a magnetic field that is almost parallel to the layers ($\cos\vartheta\leqslant\gamma_0$), the electron does not have enough time to complete a full period in the section of the fluted cylinder by the plane $p_H=\text{const}$, so that the asymptotic behavior of the components of the matrix $\tilde{\sigma}_{\alpha\beta}(k)$ changes significantly. As $\vartheta\to\pi/2$, the coupling of linearly polarized waves with the electric fields $E_z(x)$ and $E_y(x)$ weakens. The electric field $E_y(x)$ decays over distances of order δ_0, while the penetration depth for the field $E_z(x)$ increases with the magnetic field strength first in proportion to $H^{1/2}$, when $\gamma_0\geqslant\eta^{1/2}$,

$$\delta_z\simeq\frac{\delta_0}{\eta\gamma_0^{1/2}}, \tag{36}$$

and only in fairly strong fields satisfying $\gamma_0\ll\eta^{1/2}$ does it increase linearly with the magnetic field strength:

$$\delta_z\simeq\frac{\delta_0}{\eta^{3/4}\gamma_0}. \tag{37}$$

When there is an electromagnetic wave propagating along the y axis, the presence of a group of charge carriers with a quasi-one-dimensional energy spectrum does not change the penetration depth for the electromagnetic wave appreciably. As in the case of only one group of charge carriers with the dispersion law (1), the electric field along the layers decays over distances of order δ_0, while for the field perpendicular to the layers penetrates the quasi-two-dimensional conductor to a depth[25–27]

$$\delta_z\simeq\frac{\delta_0}{\eta}\Phi(\vartheta,\eta,\gamma_0), \tag{38}$$

where the function Φ strongly depends on the orientation of the magnetic field in relation to the layers. When $\cos\vartheta\leqslant\gamma_0$ holds, the penetration depth for the electric field along the normal to the layers is described by (36) and (37). But for $\cos\vartheta\geqslant\gamma_0$ the function Φ is equal to unity almost everywhere and $\delta_z\simeq\delta_0/\eta$, and only at $\vartheta=\vartheta_c$ do we have

$$\Phi=\sqrt{\frac{\cos^2\vartheta_c+(l\eta\gamma_0/\delta_0)^2}{\eta^2\cos^2\vartheta_c+\gamma_0^2}}. \tag{39}$$

Thus, by studying the magnetic-field dependence of the surface impedance when an electromagnetic wave travels in two different directions along the layers we can unambiguously determine the presence of a quasi-one-dimensional hollow in the Fermi surface and its contribution to the electrical conductivity of an organic conductor.

The above formulas were derived under the assumption $\omega\tau\ll 1$. In all the organic conductors synthesized so far this condition is sure to be met in the radio- and microwave-frequency ranges for electromagnetic waves. However, in the submillimeter range the frequency of the electromagnetic wave at low temperatures may be comparable to the collision rate of the charge carriers and the carrier mean free path may be much larger than the penetration depth. In this range the interaction of the conduction electrons and the electromagnetic field are resonant when the wave frequency ω is equal to an integral multiple of, Ω, the frequency of electron revolution in the magnetic field, and the surface impedance is sensitive to the state of the conductor surface.

Under anomalous skin effect conditions, the reflection of charge carriers that effectively interact with the electromagnetic wave from the sample boundary is close to specular,[28] so that the integral term in (19) begins to play an important role. Simple calculations lead to the following expression for the penetration depth for the electric field $E_y(x)$ with $w\ll l^{3/2}r^{1/2}\delta_0^{-2}$:

$$\delta_y=\delta_0^{4/5}r^{1/5}\left(\frac{\gamma+w}{\gamma}\right)^{2/5}, \tag{40}$$

where w is the width of the scattering indicatrix for the charge carriers scattered by the sample boundary, which in the case of a slightly bumpy surface is determined by the binary correlation function of the bumps on the surface at different points, i.e., the mean-square height of the bumps and the average length of the flat sections of the sample surface.[29] As usual, unimportant numerical factors of order unity have been omitted.

When $\eta \ll w\delta_0^2 l^{-3/2} r^{-1/2} \leqslant 1$, the penetration depth δ_z for the electric field $E_z(x)$ is of order δ_0/η, but when $w \ll l^{3/2} r^{1/2} (\eta/\delta_0)^2$, we have the following universal relationship:

$$\delta_z = \eta^{-4/5} \delta_y . \tag{41}$$

The study of the high-frequency properties of thin conductors (whose thickness is less than or of order the carrier mean free path) makes it possible to extract detailed information about the spectrum of the conduction electrons and the nature of their interaction with the conductor's surface. In particular, organic conductors are an extremely convenient object for studying the electron transfer of an electromagnetic field into the bulk of the sample in the form of slowly decaying waves and Azbel' bursts of a high-frequency field, since low-dimensional conductors possess anomalous transparency.[30]

The work was partially supported by a grant from the Ukrainian Ministry of Science (Grant 2.4/192)

*)E-mail: peschansky@ilt.kharkov.ua

[1] M. V. Kartsovnik, V. N. Laukhin, V. I. Nizhankovskiĭ, and A. A. Ignat'ev, JETP Lett. **47**, 363 (1988).
[2] M. V. Kartsovnik, P. A. Kononovich, V. N. Laukhin, and I. F. Shchegolev, JETP Lett. **48**, 541 (1988).
[3] N. Toyota, T. Sasaki, K. Murata, Y. Honda, M. Tokumoto, H. Bando, N. Kinoshita, H. Anzai, T. Ishiguro, and Y. Muto, J. Phys. Soc. Jpn. **57**, 2616 (1988).
[4] W. Kang, G. Montambaux, J. R. Cooper, D. Jerome, P. Batail, and C. Lenoir, Phys. Rev. Lett. **62**, 2559 (1989).
[5] M. V. Kartsovnik, P. A. Kononovich, V. N. Laukhin, S. I. Pesotskiĭ, and I. F. Shchegolev, Zh. Éksp. Teor. Fiz. **97**, 1305 (1990) [Sov. Phys. JETP **70**, 735 (1990)].
[6] R. Yagi, Y. Iye, T. Osada, and S. Kagoshima, J. Phys. Soc. Jpn. **59**, 3069 (1990).
[7] T. Sasaki and N. Toyota, Solid State Commun. **75**, 93 (1990).
[8] T. Osada, R. Yagi, A. Kawasumi, and S. Kagoshima, N. Miura, M. Oshima, and G. Saito, Phys. Rev. B **41**, 5428 (1990).
[9] M. V. Kartsovnik, A. E. Kovalev, V. N. Laukhin, S. I. Pesotskiĭ, and N. D. Kushch, JETP Lett. **55**, 339 (1992).
[10] N. D. Kushch, L. I. Buravov, M. V. Kartsovnik *et al.*, Synth. Met. **46**, 271 (1992).
[11] M. V. Kartsovnik, A. E. Kovalev, V. N. Laukhin, and S. I. Pesotskii, J. Phys. I France **2**, 223 (1992).
[12] A. E. Kovalev, M. V. Kartsovnik, and N. D. Kurshch, Solid State Commun. **87**, 705 (1993).
[13] A. E. Kovalev, M. V. Kartsovnik, R. P. Shibaeva *et al.*, Solid State Commun. **89**, 575 (1994).
[14] M. V. Kartsovnik, A. E. Kovalev, R. P. Shibaeva *et al.*, Physica B **201**, 459 (1994).
[15] M. V. Kartsovnik, A. E. Kovalev, V. N. Laukhin *et al.*, Synth. Met. **70**, 811 (1995).
[16] R. Rossenau, M. L. Doublet, E. Canadell *et al.*, J. Phys. I France **6**, 1527 (1996).
[17] T. Sasaki, H. Ozawa, H. Mori *et al.*, J. Phys. Soc. Jpn. **65**, 213 (1996).
[18] J. H. Condon, Phys. Rev. **145**, 526 (1966).
[19] I. A. Privorotskiĭ, Zh. Éksp. Teor. Fiz. **52**, 175 (1967) [Sov. Phys. JETP **25**, 112 (1967)].
[20] V. G. Peschanskiĭ and D. I. Stepanenko, Zh. Éksp. Teor. Fiz. **112**, 1841 (1997) [JETP **85**, 1007 (1997)].
[21] V. G. Peschanskiĭ, V. M. Kardenas, M. A. Lur'e, and K. Yiasemides, Zh. Éksp. Teor. Fiz. **80**, 1645 (1981) [Sov. Phys. JETP **53**, 849 (1981)].
[22] M. Ya. Azbel', Zh. Éksp. Teor. Fiz. **39**, 400 (1960) [Sov. Phys. JETP **12**, 283 (1961)].
[23] V. G. Peschansky, in *Soviet Scientific Reviews, Section A*, I. M. Khalatnikov (ed.), Vol. 16, 1 (1992).
[24] G. E. H. Reuter and E. H. Sondheimer, Proc. R. Soc. London **195**, 336 (1948).
[25] V. G. Peschanskiĭ, S. N. Savel'eva, and Kh. Kkheir Bek, Fiz. Tverd. Tela (Leningrad) **34**, 1630 (1992) [Sov. Phys. Solid State **34**, 864 (1992)].
[26] V. G. Peschanskiĭ, Kh. Kkheir Bek, and S. N Savel'eva, Fiz. Nizk. Temp. **18**, 1012 (1992) [Sov. J. Low Temp. Phys. **18**, 711 (1992)].
[27] V. G. Peschansky, Phys. Rep. **288**, 305 (1997).
[28] B. É. Meerovich, Zh. Éksp. Teor. Fiz. **58**, 324 (1970) [Sov. Phys. JETP **31**, 175 (1970)].
[29] L. A. Fal'kovskiĭ, Zh. Éksp. Teor. Fiz. **58**, 1830 (1970) [Sov. Phys. JETP **31**, 981 (1970)].
[30] V. G. Peschanskiĭ, Kh. A. Roldan Lopes, and D. A. Toryanik, Fiz. Nizk. Temp. **23**, 784 (1997) [Low Temp. Phys. **23**, 591 (1997)].

Translated by Eugene Yankovsky

Current-voltage characteristics of SIS structures with localized states in the material of the barrier layer

I. A. Devyatov*) and M. Yu. Kupriyanov

D. V. Skobel'tsyn Scientific-Research Institute of Nuclear Physics, M. V. Lomonosov State University, 119899 Moscow, Russia
(Submitted 20 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 687–699 (August 1998)

Elastic resonant tunneling through a single localized state in an insulating layer (I-layer) situated in the constriction zone between two thick superconducting electrodes is investigated theoretically, and the current-voltage characteristic (IVC) of the structure is calculated. The accompanying analysis leads to the prediction that an appreciable current can flow through the structure, not at $|eV|=2\Delta$ (Δ is the modulus of the order parameter of the superconducting electrodes) as in the case of an ordinary SIS junction, but at $|eV|\geqslant\Delta$, and also that the IVC can acquire segments of negative differential resistance in the case of tunneling through a single localized state. Averaging of the IVC over an ensemble of localized states distributed uniformly throughout the volume of the I-layer and with respect to the energy near the chemical potential μ in the limit $\Gamma_0/\Delta\gg 1$ (Γ_0 is the half-width of the resonance line of the localized state) produces a smaller excess current than in a junction of the SNS type. It is shown that the IVC's exhibit a transition from an excess current to a deficit current as Γ_0 decreases in the high-voltage range. © *1998 American Institute of Physics.* [S1063-7761(98)02408-1]

A great many experimental results have been published lately[1–5] in support of the hypothesis[6] that current transport in high-T_c superconducting Josephson junctions with oxide semiconductor barrier layers is a resonance phenomenon. An investigation of the stationary properties of such junctions[7] has shown that the presence of localized states in the barrier layer results in the formation of rather unique Josephson junctions of the ScS type. In contrast to the thoroughly studied ''geometrical'' ScS constrictions,[8–13] the properties of the junction type discussed below are determined not so much by geometrical factors as by the relation between the characteristic energies of the problem: the half-width of the resonance line Γ_0 and the modulus of the order parameter Δ of the superconducting electrodes.

The problem of calculating the total current through ScS structures whose transparency D ($0<D\leqslant 1$) does not depend on the quasiparticle energy has already been solved in general form.[9] Only recently,[10] however, a generalized Blonder–Tinkham–Klapwijk trajectory technique[11] has been used to carry through the calculations to specific results suitable for comparison with experiment, and it has been proved[12] that calculations based on a quasiclassical trajectory model[10] and the Keldysh Green's function formalism[9] are equivalent.

The objective of the present study is to generalize the approach developed in Ref. 10 to the case of ScS structures whose transparency D depends on the quasiparticle energy as a result of localized states present in the constriction zone. Unlike the previously investigated[14,15] phenomenon of tunneling along a ''resonant-percolation'' path consisting of a set of localized states, the present article is confined strictly to processes of single resonant tunneling of the superconducting-electrode/localized-state/superconducting-electrode type. The current transport through the structure is determined either by the average current through the set of localized states or by the current through a single localized state, depending on the density of localized states. These two limits will be systematically discussed below.

1. THE JUNCTION MODEL

We assume that the geometrical width of the constriction in the direction of the current is much smaller than the coherence length ξ of the barrier layer and the elastic and inelastic mean free paths of electrons. To simplify the calculations, we confine them to the one-dimensional model; in addition, we assume that the interface between the superconducting electrodes and the material of the constriction zone is transparent and that the dielectric barrier is in the shape of a rectangle of height W and thickness $2d$, which is localized inside the constriction and contains a negative delta-function potential $-B\delta(z-z_0)$ describing a localized state situated at a distance z_0 from the center of the rectangular potential.[16] In addition, following the same line of reasoning as in Ref. 10, we assume that the voltage V applied to the junction is fixed and that the resistance of the insulating layer (I-layer) containing a localized state is much lower than the resistance of the constriction as a whole. The voltage drop therefore takes place outside the I-layer containing the localized state. We confine the ensuing discussion strictly to an isotropic type of coupling in the S-electrodes. However, the method developed below is naturally extended to other possible types of symmetry of the order parameter in high-T_c superconducting structures.[17]

2. BASIC RELATIONS

The assumptions made enable us to reduce the problem of calculating the current-voltage characteristic (IVC) to the problem of matching the solutions of the Bogolyubov–de Gennes equations in the constriction zone while neglecting the nongradient terms in these equations and to write these solutions as a superposition of plane waves describing the propagation of electron and hole excitations of energy E in the weak-coupling region. For example, the wave function generated by electronlike excitation of the left-hand superconductor for quasiparticles accelerated by the applied potential V has the form[10]

$$\psi_{el}=\sum_n\ [(a_{2n}A_n+J_0\delta_{n,0})\exp(ikx)+B_n \times\exp(-ikx)]\exp\{-i(E+2neV)t\},$$

$$J_n=\sqrt{1-|a_n(E)|^2}(-d<z<z_0),$$

$$\psi_h=\sum_n\ [A_n\ \exp(ikx)+a_{2n}B_n\ \exp(-ikx)] \times\exp\{-i(E+2neV)t\}\quad(-d<z<z_0),$$

$$\psi_{el}=\sum_n\ [C_n\ \exp(ikx)+a_{2n+1}D_n\ \exp(-ikx)] \times\exp\{-i(E+(2n+1)eV)t\}\quad(z_0<z<d),$$

$$\psi_h=\sum_n\ [a_{2n+1}C_n\ \exp(ikx)+D_n\ \exp(-ikx)] \times\exp\{-i(E+(2n+1)eV)t\}\quad(z_0<z<d). \tag{1}$$

Here $a_m=a(E+meV)$ is the coefficient of Andreev reflection of quasiparticles from the superconducting electrodes,[10]

$$a(E)=\frac{1}{\Delta}\begin{cases}\mathrm{sgn}(E)\Delta^2(|E|+\sqrt{E^2-\Delta^2})^{-1}, & |E|\geqslant\Delta\\ E-i\sqrt{\Delta^2-E^2}, & |E|<\Delta\end{cases}, \tag{2}$$

μ is the Fermi energy, and Δ is the modulus of the order parameter of the electrodes ($\Delta\ll\mu<W$). The relation between the coefficients characterizing the wave function on opposite sides of the localized states is determined by standard matching conditions, which yield

$$\begin{pmatrix}B_n\\ C_n\end{pmatrix}=S_{el}(E_0,E_e^n)\begin{pmatrix}a_{2n}A_n+J\delta_{n,0}\\ a_{2n+1}D_n\end{pmatrix},$$

$$\begin{pmatrix}A_n\\ D_{n-1}\end{pmatrix}=S_h(E_0,E_h^n)\begin{pmatrix}a_{2n}B_n\\ a_{2n-1}C_{n-1}\end{pmatrix}. \tag{3}$$

Here $E_e^n=E+2neV$, $E_h^n=E-2(E+neV)$, and S_{el} and $S_h=S_{el}^*$ are the scattering matrices for electrons and holes, respectively:[8,10]

$$S_{el}=\begin{pmatrix}r & t\\ t & r'\end{pmatrix},\quad r(E_0,E)=\frac{up}{dn},$$

$$t(E_0,E)=\frac{1}{dn},\quad r'(E_0,E)=-\frac{up^*}{dn}, \tag{4}$$

$$dn=dn(E_0,E)=\frac{1}{\Gamma_0}\left[\frac{2k\kappa}{k^2+\kappa^2}(E_0-E)+i\left\{\left(\frac{\kappa^2-k^2}{\kappa^2+k^2}\right)\times(E_0-E)-\left(\frac{\kappa^2+k^2}{2k\kappa}\right)\left(\frac{\Gamma_1+\Gamma_2}{2}\right)\right\}\right], \tag{5}$$

$$up=up(E_0,E)=\frac{1}{\Gamma_0}\left[\left(\frac{\Gamma_2-\Gamma_1}{2}\right)+i\left\{\left(\frac{\kappa^2-k^2}{\kappa^2+k^2}\right)\times\left(\frac{\Gamma_1+\Gamma_2}{2}\right)-(E_0-E)\right\}\right],$$

$$\Gamma_{1,2}=\Gamma_0\ \exp(\pm2\kappa z_0), \tag{6}$$

$$\Gamma_0=2(W-E)\sqrt{D_0},\quad D_0=16\,\frac{k^2\kappa^2}{(k^2+\kappa^2)^2}\ \exp(-4\kappa d),$$

$$\kappa=\sqrt{2m(W-\mu)},\quad k=\sqrt{2m\mu}. \tag{7}$$

In Eqs. (1)–(7) Γ_0 is the half-width of the resonance line at the center of the barrier of a localized state having a self-energy equal to the Fermi energy μ, D_0 is the transparency of the rectangular potential without a localized state, κ is the reciprocal radius of the localized state, $\Gamma_{1,2}$ denotes the decay rates of the resonant state in the right- and left-hand electrodes, respectively, $E_0=W-B^2/8m$ is the energy of the localized state, and m is the electron effective mass. It is readily verified that Eqs. (5)–(7) impart a Breit–Wigner form to the amplitude of quasiparticle transmission through the barrier:

$$|t(E_0,E)|^2=\frac{\Gamma_0^2}{(E_R-E)^2+(\Gamma_1+\Gamma_2)^2/4}, \tag{8}$$

where

$$E_R=E_0+\frac{(k^2-\kappa^2)}{2k\kappa}\,\frac{\Gamma_1+\Gamma_2}{2}$$

is the renormalized energy of the localized state.[16]

Eliminating the coefficients C_n and D_n from Eqs. (3), we arrive at a recursion formula for B_n and A_n (to simplify the notation we have omitted the dependence of r and t on E_0):

$$F_nB_{n+1}-C_nB_n+H_nB_{n-1}=-r(E_e^n)J_0\delta_{n,0},$$

$$F_n=\frac{a_{2n+1}a_{2n+2}t(E_e^n)t^*(E_h^{n+1})}{1-a_{2n+1}^2b_n'},\quad b_n'=\frac{r'^*(E_h^{n+1})}{r'^*(E_e^n)},$$

$$H_n=\frac{a_{2n}a_{2n-1}r(E_e^n)t^*(E_h^n)t(E_e^{n-1})}{r(E_e^{n-1})(1-a_{2n-1}^2b_{n-1}')},$$

$$G_n=1-r(E_e^n)r^*(E_h^n)a_{2n}^2-\frac{a_{2n}^2a_{2n-1}^2r(E_e^n)t^{*2}(E_h^n)}{r'^*(E_e^{n-1})(1-a_{2n-1}^2b_{n-1}')}-\frac{a_{2n+1}^2r'^*(E_e^{n+1})t^2(E_e^n)}{r(E_e^n)(1-a_{2n+1}^2b_n')}. \tag{9}$$

The solutions of Eq. (9) have the form of a continued fraction[18] (see also Refs. 10 and 13):

$$B_0=-\frac{r(E_e^0)J_0}{F_0S_1-G_0+H_0S_{-1}},\quad B_n=B_0\prod_{i=1}^{|n|}S_{|i|},$$

$$S_{n,n>0}=\frac{H_n}{G_n-F_nS_{n+1}},\quad S_{n,n>0}=\frac{F_n}{G_n-H_nS_{n+1}},$$

$$A_n=r^*(E_h^n)a_{2n}B_n+a_{2n-1}\times\frac{t^*(E_h^n)t(E_e^{n-1})(B_{n-1}-a_{2n}a_{2n-1}B_nt^*(E_h^n)/t^*(E_e^{n-1}))}{r(E_e^{n-1})(1-a_{2n-1}^2b'_{n-1})}. \tag{10}$$

The wave function generated by electronlike excitation of the right-hand superconductor for quasiparticles in the I-layer has a form similar to (1):

$$\psi_{el}=\sum_n\ [(a_{2n-1}A'_n+J_1\delta_{n,1})\exp(-ikx)+B'_n\ \exp(ikx)]\times\exp\{-i(E+(2n-1)eV)t\}\quad (z_0<z<d),$$

$$\psi_h=\sum_n\ [(A'_n\ \exp(-ikx)+a_{2n-1}B'_n\ \exp(ikx)]\times\exp\{-i(E+(2n-1)eV)t\}\quad (z_0<z<d),$$

$$\psi_{el}=\sum_n\ [C'\ \exp(-ikx)+a_{2(n-1)}D'_n\ \exp(ikx)]\times\exp\{-i(E+2(n-1)eV)t\}\quad (-d<z<z_0),$$

$$\psi_h=\sum_n\ [(a_{2(n-1)}C'_n\ \exp(-ikx)+D'_n\ \exp(ikx)]\times\exp\{-i(E+2(n-1)eV)t\}\quad (-d<z<z_0). \tag{11}$$

Performing calculations analogous to those described above, we arrive at a recursion formula for B'_n similar to (9):

$$F'_nB'_{n+1}-G'_nB'_n+H'_nB'_{n-1}=-r'(E_e^{n-1})J_1\delta_{n,1},$$

$$F'_n=\frac{a_{2n}a_{2n-1}r'(E_e^{n-1})t^*(E_h^n)t(E_e^n)}{r'(E_e^n)(1-a_{2n}^2b_n)},\quad b_n=\frac{r^*(E_h^n)}{r^*(E_e^n)},$$

$$H'_n=\frac{a_{2n-2}a_{2n-3}t(E_e^{n-1})t^*(E_h^{n-1})}{1-a_{2n-2}^2b_{n-1}},$$

$$G'_n=1-r'(E_e^{n-1})r'^*(E_h^n)a_{2n-1}^2-\frac{a_{2n}^2a_{2n-1}^2r'(E_e^{n-1})t^{*2}(E_h^n)}{r^*(E_e^n)(1-a_{2n}^2b_n)}-\frac{a_{2n-2}^2r^*(E_h^{n-1})t^2(E_e^{n-1})}{r'(E_e^{n-1})(1-a_{2n-2}^2b_{n-1})}. \tag{12}$$

Introducing the new index $s=1-n$ and the change of notation $\beta_s=B'_{1-s}=B'_n$ we can reduce the recursion formulas (12) to the form (9) with $\delta_{n,0}$ on the right-hand side. The solutions (12) therefore acquire a form similar to (10):

$$\beta_0=-\frac{r'(E_e^0)J_1}{f_0s_1-g_0+h_0s_{-1}},\quad \beta_n=\beta_0\prod_{i=1}^{|n|}s_{|i|},$$

$$s_{n,n>0}=\frac{h_n}{g_n-f_ns_{n+1}},\quad s_{n,n<0}=\frac{f_n}{g_n-h_ns_{n-1}},$$

$$\alpha_n=A'_{1-n}=r'^*(E_h^{1-n})a_{1-2n}\beta_n+t^*(E_h^{1-n})a_{2(1-n)}\times\frac{\beta_{n-1}t(E_e^{1-n})/r'(E_e^{1-n})+\beta_na_{1-2n}a_{2(1-n)}t^*(E_h^{1-n})/r^*(E_e^{1-n})}{1-a_{2(1-n)}^2b_{1-n}}, \tag{13}$$

where $h_n=F'_{1-n}$, $f_n=H'_{1-n}$, $g_n=G'_{1-n}$, and $\alpha_n=A'_{1-n}$.

It is convenient to work with the current through the junction in the constriction zone, where it is expressed by the usual quantum-mechanical formula for one-particle excitations and is represented by a Fourier series:

$$I(t)=\sum_k\ I_k\ \exp(2ikeVt),$$

$$\frac{I_k\pi\hbar}{e}=eV\delta_{k,0}+\int d\varepsilon\ f(\varepsilon)\Big\{J_0[a_{2k}^*A_k^*+a_{-2k}A_{-k}]+\sum_n\ (1+a_{2n}a_{2(n+k)}^*)(A_nA_{n+k}^*-B_nB_{n+k}^*)\Big\}-\int d\varepsilon\ f(\varepsilon+eV)\Big\{J_1[a_{2k+1}^*\alpha_{-k}^*+a_{1-2k}\alpha_k]+\sum_n\ (1+a_{2n-1}a_{2(n+k)-1}^*)(\alpha_{1-n}\alpha_{1-(n+k)}^*-\beta_{1-n}\beta_{1-(n+k)}^*)\Big\}. \tag{14}$$

Equations (10)–(14) determine all the components of the current through the structure with a localized state. If the scattering matrices S_e and S_h associated with the presence of the localized state are independent of the energy, the expression for the current (14) and the recursion formulas (10) and (13) go over to Eqs. (6) and (5) in Ref. 10, where now

$$A'_n(E)=-A_{1-n}^*(-E-eV),$$

$$B'_n(E)=(r/r^*)B_{1-n}^*(-E-eV).$$

3. CALCULATION OF THE AVERAGED CURRENT

If a large number of localized states are present in the barrier, the current through the junction is a self-averaged quantity, and Eqs. (14) must be averaged over the energies of the localized states and their spatial distribution. Assuming that the localized states are distributed uniformly with a density g in the weak-coupling zone and with respect to the energy in the vicinity of the Fermi surface, we can show that in the limit $eV\gg\Delta$ the asymptotic expression for the constant current through the structure, averaged over the coordinates and the energies of the localized states, has the form

$$I_{dc}(V)=R_N^{-1}eV+\delta I,\quad eV\gg\Delta,\quad R_N^{-1}=\pi^2g\Gamma_0/2\kappa, \tag{15}$$

where R_N^{-1} is the averaged resonant conductance of an analogous structure with normal electrodes.[19]

At low temperatures, $T\ll\Delta$, we write the term δI in (15) in the form

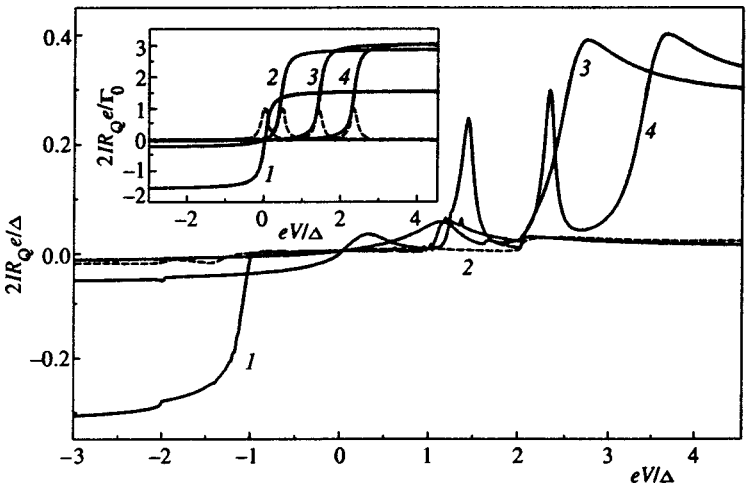


FIG. 1. IVC's of an ScS structure with a single localized state in the weak-coupling zone calculated numerically on the basis of Eqs. (9)–(14) for $\Delta/\Gamma_0=10$, $z_0=0$, and various values of the energy of the single localized state: *1)* $E_R=0$; *2)* $E_R=-0.4\Gamma_0$; *3)* $E_R=-1.4\Gamma_0$; *4)* $E_R=-2.3\Gamma_0$. The case $E_R=-0.4\Gamma_0$ is represented by a dashed curve. Inset: for comparison, IVC's for analogous NcN structures with nonsuperconducting electrodes calculated numerically for the same values of the parameter; the dashed lines represent the conductance as a function of voltage.

$$\delta I=\frac{e}{\pi\hbar}\int d\varepsilon\left\langle\frac{|t^*(\varepsilon)t(-\varepsilon)|^2}{|1-a^2(\varepsilon)r^*(-\varepsilon)r(\varepsilon)|^2}\right\rangle(|a(\varepsilon)|^2+|a(\varepsilon)|^4)+\frac{e}{\pi\hbar}\int_{|\varepsilon|>\Delta}d\varepsilon\left\langle|t(\varepsilon)|^2\times\left\langle\frac{1-|a(\varepsilon)|^4}{|1-a^2(\varepsilon)r^*(-\varepsilon)r(\varepsilon)|^2}-1\right\rangle\right\rangle-\frac{2\Delta}{eR_N},\qquad(16)$$

where the angle brackets signify averaging over the energy E_0 and the coordinate z_0 of the localized state.

In the limit of a broad resonance line, $\Gamma_0\gg\Delta$, the differences between the pair $t(\varepsilon)$, $t(-\varepsilon)$ and the pair $r(\varepsilon)$, $r(-\varepsilon)$ in the integrals over the energy (16) can be disregarded in the significant energy interval $\varepsilon\approx\Delta$, whereupon they can be reduced to tabulated integrals:

$$\delta I=\frac{e\Delta}{\pi\hbar}\left\langle T^2\left(\frac{1}{1-T}+\frac{2}{\sqrt{1-T}(2-T)}\operatorname{arctanh}\left[\frac{2\sqrt{1-T}}{2-T}\right]-\frac{2-T}{(1-T)^{3/2}}\operatorname{arctanh}[\sqrt{1-T}]\right)\right\rangle,\qquad(17)$$

where $T\equiv|t(E_0,E|^2$. Numerical averaging over the energy and coordinate of the localized state in Eq. (17) gives positive values of the excess current $\delta I\simeq 0.22(2\Delta/\varepsilon R_N)$.

In the opposite limit of a narrow resonance line, $\Delta\gg\Gamma_0$, analytical estimates and numerical calculations of the triple integrals in Eq. (16) yield negative values of δI, which correspond to a deficit current through the structure, $\delta I\simeq-0.7(2\Delta/eR_N)$.

The results concur with those in Ref. 15, where the Keldysh Green's function approach led to the prediction of a similar transition from a deficit current to an excess current in tunneling along resonant-percolation paths consisting of a large number of localized states. It should also be noted that in the limit $\Gamma_0\gg\Delta$, despite a qualitative similarity to ballistic ScS constrictions, the expression obtained for the excess current is 1.5 times smaller than the one obtained for ScS junctions.

4. CURRENT THROUGH A SINGLE LOCALIZED STATE

It is interesting to determine the current through a single localized state for a low density of localized states g in the I-layer.

Let us first consider current transport through the analogous structure with normal reservoirs. This problem has been investigated previously.[16,19] Inasmuch as Andreev reflection processes can be ignored when there is no superconductivity in the electrodes, the expression for the current (14) reduces to the equation

$$I=\frac{e}{\pi\hbar}\int d\varepsilon\{f(\varepsilon)|t(E_R,\varepsilon)|^2-f(\varepsilon+eV)|t(E_R,\varepsilon)|^2\}\xrightarrow{T\to 0}\frac{e}{\pi\hbar}\times\int_0^{eV}d\varepsilon|t(-E_R,\varepsilon)|^2.\qquad(18)$$

It follows from Eq. (18) that at low temperature, $T\ll\Gamma_0$, the derivative of the current $dI(V)/dV$ takes the form of a Lorentzian plot of $|t(-E_R,V)|^2$ with a maximum at $eV=-E_R$, and the current $I(V)$ is represented by a step, which is diffuse on the scale of Γ_0, beginning its rise at $eV\ll-E_R$ and leveling off at a constant amplitude $(e/\pi\hbar)\times[\pi\Gamma_0/\cosh(z_0)]$ at $eV\gg-E_R$. This conclusion is valid for $E_R\ll-\Gamma_0$. For $E_R\gg 0$ the resonant current is insignificant at $V>0$, and for $E_R=0$ it corresponds to a linear segment at $V=0$, exhibits a linear dependence as long as $eV\ll\Gamma_0$, and approaches a constant value $(e/\pi\hbar)[\pi\Gamma_0/2\cosh(z_0)]$ at $eV\gg\Gamma_0$. The IVC's calculated numerically according to Eq. (18) for a structure with normal reservoirs are shown in the inset to Fig. 1 for various values of the localized-state energy.

The presence of superconductivity in the electrodes of the structure makes it necessary to include numerous processes involving the Andreev reflection of quasiparticles from the superconducting electrodes in addition to the one-electron scattering described by the matrix (2). The current

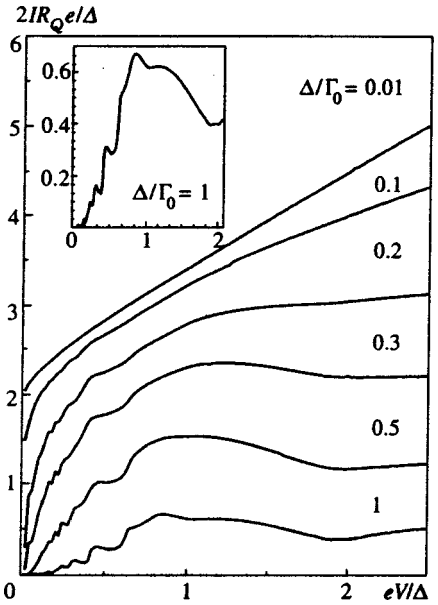


FIG. 2. IVC's of an ScS structure with a single localized state in the weak-coupling zone for various relationships between the half-width of the resonance line Γ_0 and the modulus of the order parameter Δ of the superconducting electrodes calculated numerically on the basis of Eqs. (9)–(14) for $E_R=0$ and $z_0=0$; $R_Q=h/4e^2$ is the quantum unit of resistance. Inset: IVC calculated for $\Delta/\Gamma_0=1$, $E_R=0$, and $z_0=0$ on a magnified scale.

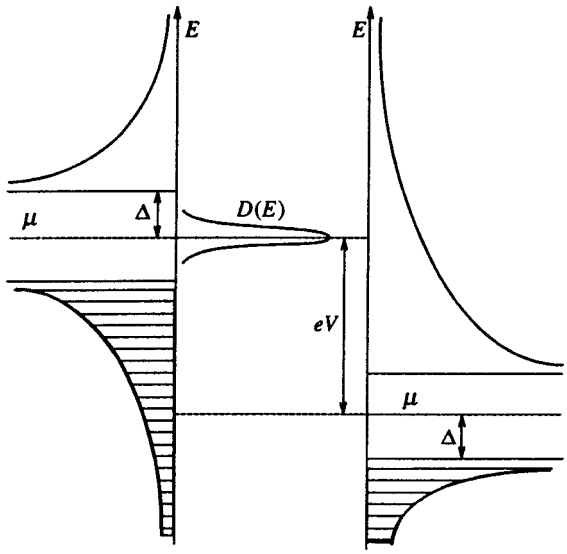


FIG. 3. ''Semiconductor'' diagram illustrating the impossibility of current flow through a structure with a localized state for a self-energy $E_R=0$ and for a junction voltage $eV>-\Delta$.

through the structure is no longer determined from Eq. (18), but from the more general expression (14). We now discuss various relations between Δ and Γ_0.

For a broad resonance line, $\Gamma_0 \gg \Delta$, and a localized-state energy $E_R \approx 0$ analytical determinations and a numerical calculation according to Eq. (14) yield an IVC having a form similar to the IVC of a ballistic SNS constriction with transparency $D=1$ (Fig. 2). This result is entirely natural: in this case quasiparticles move freely in the constriction zone in their significant energy interval $\varepsilon \approx \Delta \ll \Gamma_0$, undergoing many Andreev reflections in the subgap region. This process induces a current jump $2\Delta e/\pi\hbar$ at $V=0$ and an excess current $\delta I = 8\Delta e/3\pi\hbar$ at $eV \gg \Delta$, just as in the SNS case.[9]

For a resonance linewidth $\Gamma_0 \approx \Delta$ and a localized-state energy close to zero, $E_R \approx 0$, a numerical calculation according to Eq. (14) imparts a complex form to the $I(V)$ curve. The IVC's acquire prominent features in the subgap region at $eV \simeq 2\Delta/n$, segments with a negative differential resistance, and current saturation in the high-voltage region (see the inset to Fig. 2). This result also admits a natural explanation: a finite probability of normal quasiparticle scattering in the constriction zone emerges for $\Gamma_0 \approx \Delta$. This effect alters the pattern of known subgap features at $eV \simeq 2\Delta/n$ (Refs. 10 and 13) typical of structures with $D \leqslant 1$, but with a different amplitude from that obtained for an energy-independent, delta-function scattering potential. The onset of saturation at $eV \gg \Gamma_0$ is typical of structures with current transport through a single localized state.

An interesting feature of the IVC of a structure in the narrow-resonance limit $\Gamma_0 \ll \Delta$ is the possibility of the appearance of a major current resonance at $|eV| \geqslant \Delta$ for definite values of the localized-state energy, along with the emergence of segments where the IVC drops abruptly (Fig. 1). This behavior is associated with the formation of an effective single coherent Andreev process in the narrow band of voltages $|eV-E_R| \approx \Gamma_0 \ll \Delta$ at definite values of the localized-state energy and a departure from this resonance as the voltage is increased.

For example, if the localized-state energy E_R coincides with the Fermi energy of the left-hand electrode (Fig. 3), the current through the structure is not very significant (curve *1* in Fig. 1) for any positive voltage V on the junction. Indeed, the Fermi energy of the left-hand superconducting electrode in this case exceeds the energy of the right-hand electrode by the amount eV. Therefore, at a low temperature $T \ll \Delta$ there are no energy levels of the right-hand electrode filled with quasiparticles opposite the Lorentzian transparency window $|t(E_R,E)|^2$ (8) created by the localized state, and thus no current flows from the right- to the left-hand electrode. Current flowing in the opposite direction is also suppressed by virtue of the inequality $\Gamma_0 \ll \Delta$.

For negative voltages on the junction, $V \approx -\Delta/e$ (see Fig. 4), there are filled levels of the right-hand superconductor opposite the Lorentzian transparency window $|t(E_R,E)|^2$ described by Eq. (8). The resulting effective single Andreev reflection of electrons moving from the right- to the left-hand electrode in the narrow band $|V-E_R| \approx \Gamma_0 \ll \Delta$ induces a significant resonant current. A similar situation has also occurred in the case of an equilibrium resonant Josephson current at $V=0$ (Ref. 7). Accordingly, in the limit of a narrow resonance line, $\Gamma_0 \ll \Delta$, once again a resonant Josephson current could flow effectively through such a structure only in the narrow energy band $E_R \approx \Gamma_0$. As the absolute value of the voltage is increased, the current smoothly approaches a constant value in a way reminiscent of the behavior in a structure with normal electrodes.

For self-energies of the localized state E_R that do not coincide with the Fermi energy of the left-hand electrode and are smaller than Δ, an effective Andreev process does not occur for any voltage on the junction. This result precludes

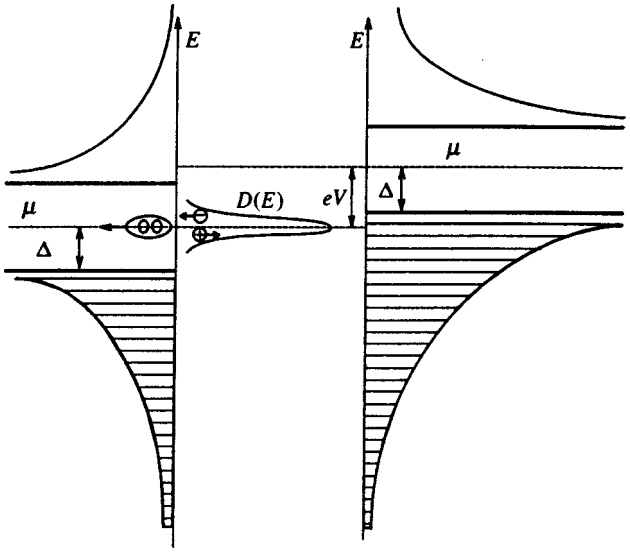


FIG. 4. ''Semiconductor'' diagram explaining the appearance of an appreciable current through a structure with a localized state for a self-energy $E_R=0$ and for a junction voltage $eV<-\Delta$ as a result of the occurrence of an effective Andreev process relative to the Fermi level of the left-hand superconductor in the narrow energy band for transparency of these states.

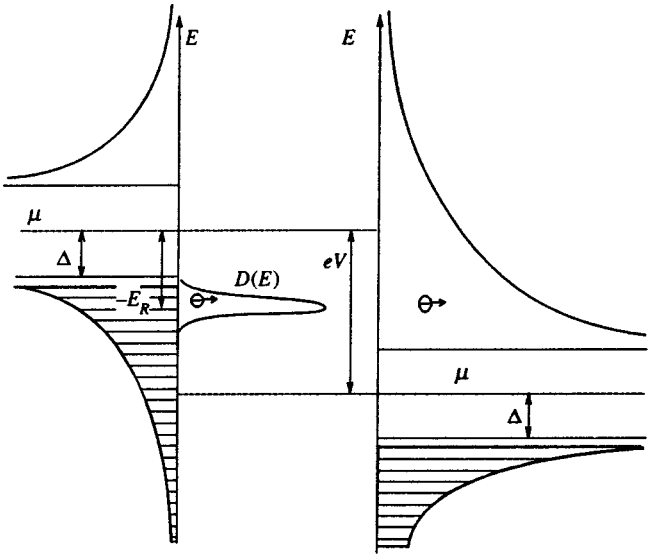


FIG. 6. ''Semiconductor'' diagram illustrating current transport through a structure with localized states for a self-energy $E_R<-\Delta$ and a junction voltage $eV>\Delta+|E_R|$ as a result of one-electron transport in the narrow energy band for transparency of these states.

any appreciable current through the structure (curve *2* in Fig. 1).

For $|E_R|>\Delta$, on the other hand, it is once again possible for a resonant current to flow through the structure (curves *3* and *4* in Fig. 1). Now the most prominent feature of the IVC of the structure is the presence of a very pronounced descending segment. At $eV\simeq -E_R$ the current rises abruptly in connection with the emergence of an effective Andreev process in the narrow energy band $|eV-E_R|\approx\Gamma_0\ll\Delta$ (see Fig. 5). A further increase in the voltage ($-E_R<eV<\Delta+|E_R|$) leads to a disparity between the transparencies for electron and hole excitations, which suppresses the Andreev process. At $eV>\Delta+|E_R|$, however, a channel of one-electron resonant current transport opens up in much the same way as in resonant structures with normal electrodes (Fig. 6). A further increase in the voltage causes the current to achieve a constant value, which tends to the analogous constant in NcN junctions as the localized-state energy decreases, $E_R\ll-\Delta$.

An increase in the temperature of the structure and shifting of the localized-state coordinate z_0 from the optimum $z_0=0$ lower the amplitude and spread out the features of the IVC of the structure.

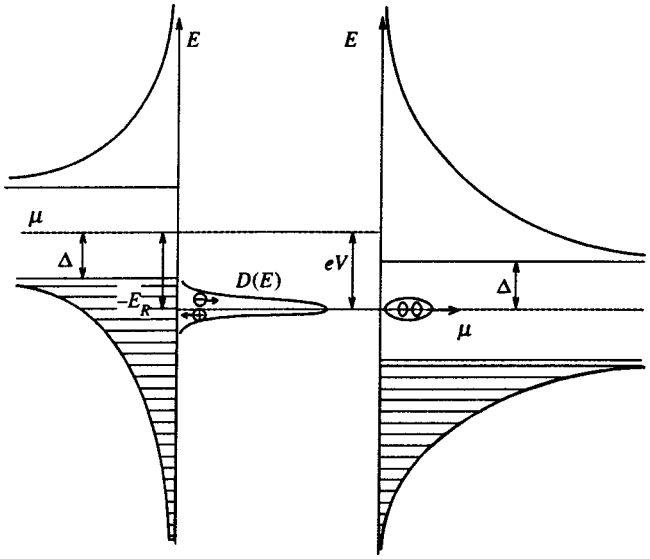


FIG. 5. ''Semiconductor'' diagram illustrating current transport through a structure with localized states for a self-energy $E_R<-\Delta$ and a junction voltage $eV>\Delta$ as a result of the occurrence of an effective Andreev process relative to the Fermi level of the left-hand superconductor in the narrow energy band for transparency of these states.

5. CONCLUSION

To summarize, we have shown that the IVC's of SIS structures having a localized state in the material of the barrier layer depend strongly on the parameter Δ/Γ_0 and on the density of localized states. If their density is low, the current is determined by transport through a single localized state. In this case, for a broad resonance line, $\Gamma_0\gg\Delta$, the IVC of the junction coincides with the IVC of a ballistic junction having a transparency $D=1$. For intermediate resonance linewidths, $\Gamma_0\approx\Delta$, the IVC of the junction exhibits a modified pattern of known subgap features[10,13] at $eV\simeq 2\Delta/n$ with the current subsequently approaching saturation. In the limit of a narrow resonance line, $\Gamma_0\ll\Delta$, at certain values of the localized-state self-energy E_R an appreciable current can flow through the junction when it is subjected to an applied voltage $|eV|=\Delta$. Segments characterized by a negative differential resistance can also be observed on the IVC of the junction.

For a high density of localized states the current through the junction is not determined by the current through a single state, but by the average contribution from a set of such states distributed uniformly throughout the volume of the I-layer and with respect to the energy near the Fermi energy. Now in the high-voltage limit $eV\gg\Delta$ an excess current

should be observed on the IVC of the junction for a broad resonance line, $\Gamma_0 \gg \Delta$, and a deficit current should be observed in the limit of a narrow resonance line, $\Gamma_0 \ll \Delta$. This result is consistent with previous results[15] obtained by a different technique for structures qualitatively similar to those discussed in this paper. We have calculated numerically the values of the deficit and excess currents.

Our investigation has focused on the case of ordinary isotropic S-type superconductors. However, the proposed method can be extended in a natural way to the case of *d* symmetry.[17]

The authors are grateful to D. Goncharov for assisting with the numerical calculations. This work has received support from the programs ''Critical Problems in Physics of the Condensed State'' and ''Physics of Solid-State Nanostructures,'' as well as jointly from the Russian Fund for Fundamental Research and the International Association for the Promotion of Cooperation with Scientists from the New Independent States of the Former Soviet Union (RFFI-INTAS Grant RFBR95-1305).

*)E-mail: idev@rsfq.npi.msu.su

[1] A. A. Golubov, M. A. J. Verthoeven, I. A. Devyatov *et al.*, Physica C (Amsterdam) **235–240**, 1361 (1994).
[2] I. I. Vengrus, M. Yu. Kupriyanov, O. V. Snigirev *et al.*, JETP Lett. **60**, 381 (1994).
[3] T. Satoh, M. Hidaka, M. Yu. Kupriyanov *et al.*, IEEE Trans. Appl. Supercond. **5**, 2612 (1995).
[4] M. Yu. Kupriyanov and J. S. Tsai, IEEE Trans. Appl. Supercond. **5**, 2531 (1995).
[5] M. Siegel, R. Dommel, C. Horstmann, and A. I. Braginskii, in *Extended Abstracts of International Conference on Superconductor Electronics* (Nagoya, Japan, 1995), p. 141.
[6] M. Yu. Kupriyanov, J. Low Temp. Phys. **106**, 149 (1997).
[7] I. A. Devyatov and M. Yu. Kupriyanov, Zh. Éksp. Teor. Fiz. **112**, 189 (1997) [JETP **85**, 189 (1997)].
[8] C. W. J. Beenakker and H. van Houten, Phys. Rev. Lett. **66**, 3056 (1991).
[9] A. B. Zaĭtsev, Zh. Éksp. Teor. Fiz. **86**, 1742 (1984) [Sov. Phys. JETP **59**, 1015 (1984)].
[10] D. Averin and A. Bardas, Phys. Rev. Lett. **75**, 1831 (1995).
[11] T. M. Klapwijk, G. E. Blonder, and M. Tinkham, Physica B (Amsterdam) **109/110**, 1657 (1982).
[12] A. V. Zaĭtsev and D. V. Averin, e-print: cond-mat/9708190 (1997).
[13] E. N. Bratus, V. S. Shumeiko, and G. Wendin, Phys. Rev. Lett. **74**, 2110 (1995).
[14] L. S. Aslamazov and M. V. Fistul', Zh. Éksp. Teor. Fiz. **83**, 1170 (1982) [Sov. Phys. JETP **56**, 666 (1982)].
[15] A. V. Tartakovskiĭ and M. V. Fistul', Zh. Éksp. Teor. Fiz. **94**(9), 353 (1988) [Sov. Phys. JETP **67**, 1935 (1988)].
[16] H. Knauer, J. Richter, and P. Seidel, Phys. Status Solidi A **44**, 303 (1977).
[17] M. Hurd, e-print: cond-mat/9702028 (1997).
[18] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[19] A. I. Larkin and K. A. Matveev, Zh. Éksp. Teor. Fiz. **93**, 1030 (1987) [Sov. Phys. JETP **66**, 580 (1987)].

Translated by James S. Wood
Edited by P. Shelnitz

Linear and nonlinear excitonic absorption in semiconducting quantum wires crystallized in a dielectric matrix

V. S. Dneprovskiĭ and E. A. Zhukov

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

E. A. Muljarov and S. G. Tikhodeev*[)]

Institute of General Physics, Russian Academy of Sciences, 117842 Moscow, Russia
(Submitted 21 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 700–710 (August 1998)

Spectra of linear and nonlinear absorption of GaAs and CdSe semiconducting quantum wires crystallized in a transparent dielectric matrix (inside chrysotile-asbestos nanotubes) have been measured. Their features are interpreted in terms of excitonic transitions and filling of the exciton phase space in the quantum wires. The theoretical model presented here has allowed us to calculate the energies of excitonic transitions that are in qualitative agreement with experimental data. The calculated exciton binding energies in quantum wires are a factor of several tens higher than in bulk semiconductors. The cause of this increase in the exciton binding energy is not only the size quantization, but also the ''dielectric enhancement,'' i.e., stronger attraction between electrons and holes owing to the large difference between permittivities of the semiconductor and dielectric matrix. © *1998 American Institute of Physics.* [S1063-7761(98)02508-6]

1. INTRODUCTION

In recent years, research attention has been focussed on quantum wires, i.e., semiconducting structures where current carriers can move freely in one direction, because they not only demonstrate interesting properties, but also show promise in view of their application to electronic and optoelectronic devices. As a result of a transition from a two-dimensional system with its stepped density of states to a one-dimensional system, we have narrow peaks in the density of electronic states. Additional quantum confinement gives rise a narrower gain spectrum, hence to a higher differential gain,[1] to a higher binding energy of excitons and a more intense laser generation on the excitonic transition,[2] and to stronger optical nonlinearities.[3] Thus, using quantum wires, one can upgrade the parameters of lasers (by obtaining a lower generation threshold and broader modulation frequency band, and limiting the effect of temperature on laser characteristics), transistors (by increasing the carrier mobility), and optical switches (by reducing both the switching energy and relaxation time).

This paper presents measurements of linear absorption in GaAs and CdSe quantum wires crystallized in a transparent matrix and nonlinear absorption subject to powerful picosecond laser pulses. The features observed in linear absorption spectra can be interpreted in terms of excitonic transitions. Energies of excitonic transitions calculated using a variational procedure (with due account of image potentials and size quantization of both electrons and holes) are in accordance with the measurements. Owing to the effect of dielectric strengthening,[4,5] the exciton binding energies are very high (more than 100 meV). Physical processes leading to nonlinear effects in the absorption have been analyzed, in particular, filling of the exciton phase space, screening of excitons, and renormalization and filling of one-dimensional electron and hole bands when high-density plasma is generated in semiconductors.

2. EXPERIMENTAL TECHNIQUES AND RESULTS

There are various techniques for fabrication of semiconducting quantum wires, namely, the molecular-beam epitaxy (MBE) or metal-organic chemical vapor-phase deposition on preprocessed substrates, etching of two-dimensional semiconducting structures, and cleaving of a two-dimensional structure in a plane perpendicular to the surface with the MBE process continued on the cleaved surface (fabrication of a T-shaped quantum wire), etc.[6] These techniques, however, do not produce samples in which the dimensions and density of quantum wires would allow the researchers to measure linear and nonlinear absorption spectra without an optical near-field microscope. We used samples manufactured using an alternative technique:[7] melted semiconducting material was injected in hollow nanometer channels of chrysotile-asbestos tubes and crystallized. Given the large sample dimensions and density of crystalline quantum wires, we could measure both linear and nonlinear absorption. The samples were densely packed regular structures of parallel chrysotile-asbestos nanotubes with an external diameter of about 30 nm containing GaAs and CdSe crystalline wires. The internal diameters of chrysotile-asbestos nanotubes were measured by a high-resolution electron microscope.[1)] The samples in which GaAs and CdSe wires were grown contained nanotubes of two types: most of them had an inside diameter of 4.8 nm, the rest had a diameter of about 6 nm.

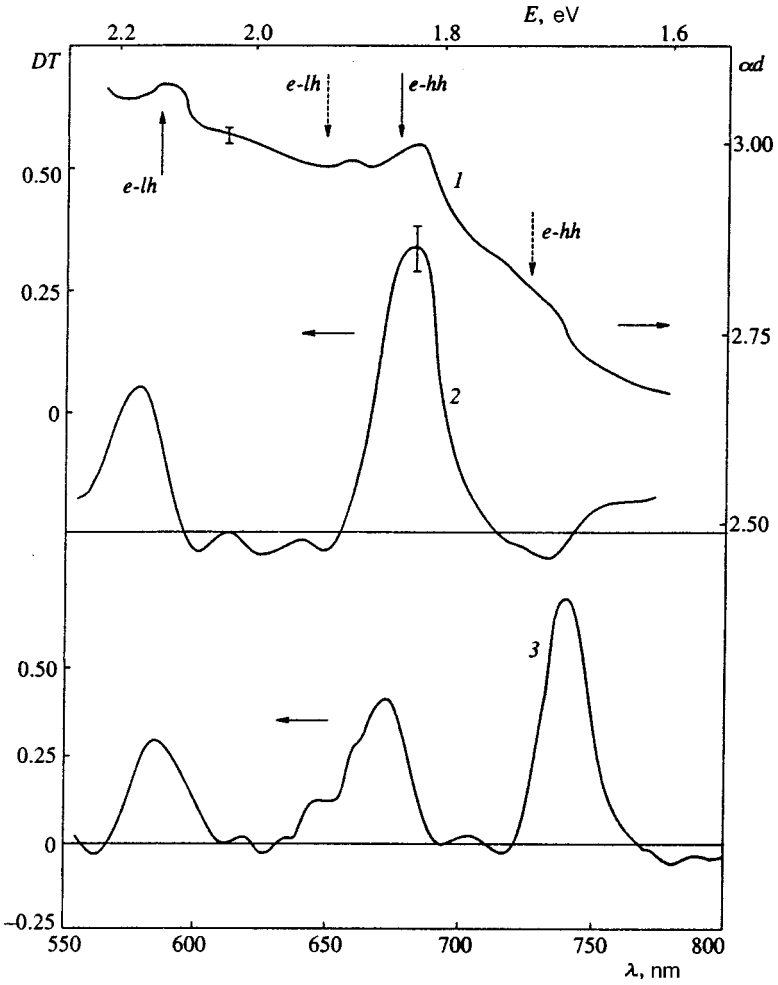


FIG. 1. Spectra of (*1*) linear absorption and (*2, 3*) differential transmission at zero delay between the pumping and probing pulses for different regions of the sample of GaAs quantum wires crystallized inside chrysotile-asbestos nanotubes. The solid and dashed arrows show calculations of excitonic transition energies in quantum wires with diameters of 4.8 nm and 6 nm, respectively.

The quantum wires were excited by ultrashort optical pulses (with a duration of about 20 ps) of the second harmonic of a Nd:YAG mode-locked laser ($\hbar\omega = 2.33$ eV). The pump intensity was up to 100 MW/cm^2. The pumping laser beam polarized in the direction parallel to the quantum wires was incident normally on the sample surface and focused into a spot with a diameter of 200 μm. The central part of the excited spot was probed by a focused ''white'' light. In order to generate the ultrashort white probing pulse, a fraction of laser light at the fundamental frequency was fed to a cell filled with heavy water. An optical delay line allowed us to delay the probing pulse from the pumping laser pulse and study the kinetics of changes induced by the pumping radiation. The spectrum of the probing pulse (''picosecond continuum'') was recorded both before and behind the sample using an OVA-284 optical multichannel analyzer.

In our experiments, we measured the differential transmission

$$DT(\lambda) = \frac{T(\lambda) - T_0(\lambda)}{T_0(\lambda)}, \tag{1}$$

where $T(\lambda)$ and $T_0(\lambda)$ are the transmission spectra of the excited and nonexcited sample.

The spectra of linear absorption and differential transmission of GaAs quantum wires crystallized inside chrysotile-asbestos nanotubes are given in Fig. 1. Note the features in the linear absorption spectrum (curve *1*), namely the broad lines peaking at 1.82 eV, 1.89 eV, and 2.11 eV, alongside the ''shoulder'' at 1.69 eV against the background of absorption increasing with decreasing wavelength. The positions of bleaching bands in the differential transmission spectrum coincided with the broad bands and the shoulder in the linear absorption spectrum. The light-induced bleaching bands at 1.84 eV and 2.14 eV (curve *2*) disappeared after 50 ps. The shorter lifetime of the 1.68-eV band could not be measured because of the insufficient time resolution of our facility. Note that the low-frequency band at 1.68 eV was detected (curve *3*) only in some regions of the samples (when both pumping and probing beams were scanned across sample surfaces). The differential transmission spectra of most samples were similar to the curve *2* in Fig. 1.

The linear absorption and differential transmission spectra of samples containing CdSe quantum wires crystallized inside chrysolite-asbestos nanotubes are given in Fig. 2. Against the background of absorption increasing with decreasing wavelength, one can see two broad features: one short-wave band at 1.98 ± 0.08 eV and one long-wave band about 1.8 eV. The positions of induced bleaching bands in the spectrum of differential transmission with zero delay between the pumping and probing pulses coincide with the features in the linear absorption spectrum.

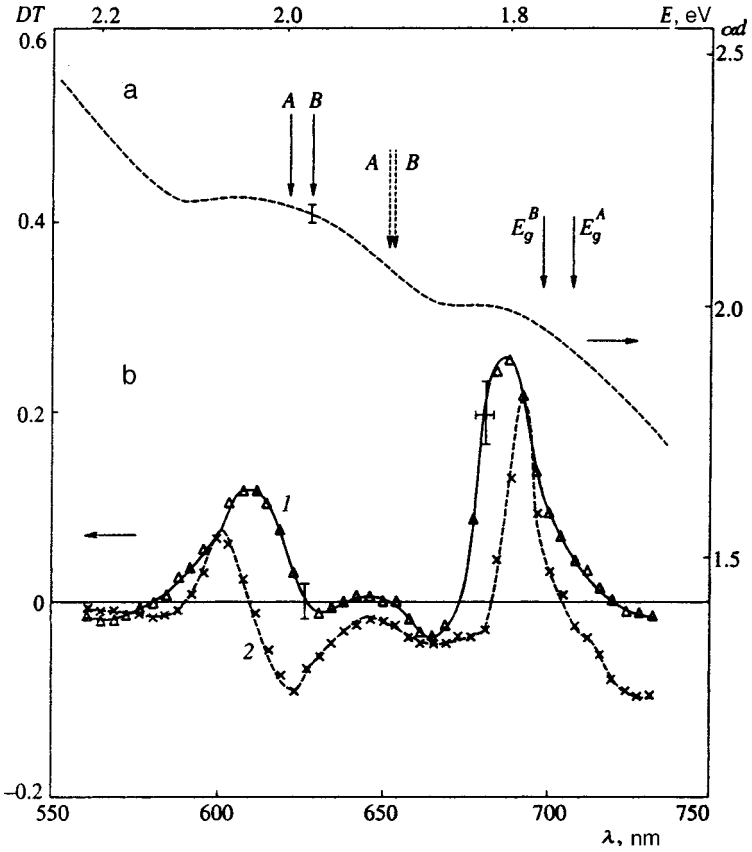


FIG. 2. Spectra of (a) linear absorption and (b) differential transmission [(*1*) at zero delay and (*2*) 7-ps delay between the pumping and probing pulses] of CdSe quantum wires crystallized inside chrysotile-asbestos nanotubes. The solid and dashed arrows show calculations of excitonic transition energies for quantum wires with diameters of 4.8 nm and 6 nm, respectively.

3. THEORETICAL MODEL

Our theoretical description of the features in the linear absorption spectrum of quantum wires is based on the model of a Wannier exciton localized in a thin cylindrical semiconducting wire inside a dielectric matrix.[8] Given the large difference (a factor of five to six) between the permittivities of the semiconductor and matrix, image potentials play an important role, alongside the size quantization, which determines the quasi-one-dimensional character of excitons. These potentials lead to the effect of dielectric strengthening[4,5] (the exciton binding energy increases by a factor of several tens) and to renormalization of localizing potentials in quantum wires owing to additional self-action potentials.[9]

The separation between size quantization levels in GaAs (CdSe) quantum wires varies between 100 meV for heavy holes (*A* and *B* valence subbands) to several hundreds of meV for light holes and electrons. These values are comparable to or higher than the calculated exciton binding energy (see Table I) and much larger than the Coulomb correlation energy between size quantization levels. This allows us to separate the carrier motion in the localizing potential perpendicular to the wire axis from the relative electron–hole motion along the quantum wire axis. The complex spectrum of the valence band in GaAs and CdSe and the hybridization of the light and heavy hole subbands (*A* and *B* subbands) in the localizing potential deserve a dedicated investigation (see, for example, Ref. 10). Under the conditions of size quantization, however, the valence band degeneracy is lifted (a quasi-degeneracy takes place), and the hole subbands are split from one another (their effective masses remain anisotropic). The effects of band nonparabolicity in this case are suppressed, which allows us to neglect the hole mass renormalization when the subband splitting is large and use the hole effective masses[11] obtained at room temperature. Thus, our theoretical model considers two types of excitons. In GaAs we have $e-hh$ excitons formed by holes which are light along the wire axis and heavy in the plane where the motion is confined, and $e-lh$ excitons with holes which are heavy along the wire axis and light in the confinement plane. In CdSe, these are excitons formed by holes from the *A* and *B* valence subbands.

Note that the two-dimensional localizing potentials acting on electrons and holes include, as in Ref. 9, the self-action potentials of the charges.[2)] The self-action potentials greatly modify the quantum well shape by reducing its depth, which affects primarily the renormalized semiconductor band gap, hence the exciton peak position.

The exciton parameters, namely the binding energy, wave function, and mean linear size along the wire axis, are calculated using the variational technique[8] taking into account the contribution of image potentials to the energy of the Coulomb interaction between electrons and holes averaged over wave functions of transverse motion of carriers.

4. DISCUSSION OF RESULTS

Unlike three- and two-dimensional semiconducting systems, optical properties of quasi-one-dimensional systems are largely controlled by excitonic transitions owing to the anomalously large concentration of the oscillator strength at

TABLE I. Calculations of excitonic parameters of GaAs and CdSe quantum wires crystallized in chrysotile-asbestos nanotubes and measurements of excitonic transition energies (the notations are explained in the text).

Wire material and exciton type	Model parameters[10]						Theory					Experiment
	$\frac{m_{\perp e}}{m_0}$	$\frac{m_{\perp h}}{m_0}$	$\frac{\mu}{m_0}$	E_g^0, eV	ε_∞^w	ε_∞^b	d, nm	E_b, meV	E_g, eV	L_{exc}, nm	E_{exc}, eV	E_{exc}, eV
GaAs $e-hh$	0.067	0.50	0.034	1.426	10.9	2.2	4.8	160	2.000	7.5	1.840	1.82±0.04
GaAs $e-lh$	0.067	0.068	0.059	1.426	10.9	2.2	6.0	133	1.843	8.8	1.710	1.69±0.04
							4.8	165	2.284	6.9	2.118	2.11±0.05
CdSe A	0.12	0.45	0.107	1.751	5.8	2.2	6.0	137	2.047	8.2	1.910	1.89±0.05
							4.8	196	2.188	5.6	1.992	
CdSe B	0.12	0.9	0.107	1.771	5.8	2.2	6.0	161	2.066	6.8	1.905	1.98±0.08
							4.8	198	2.173	5.6	1.975	
							6.0	163	2.063	6.7	1.900	

frequencies of excitonic transitions.[12,13] In linear absorption spectra of chrysotile-asbestos samples containing GaAs (Fig. 1, curve *1*) and CdSe (Fig. 2a), we also observe absorption which grows monotonically with the photon energy. It seems that a fraction of the superconducting material is crystallized between bunches of chrysotile-asbestos nanotubes in the form of relatively large microcrystals. The interband absorption in the bulk semiconductor (microcrystals unaffected by size quantization) can contribute a monotonic component to the absorption spectrum.

This conjecture is confirmed by the spectra of the differential transmission of chrysotile-asbestos samples containing CdSe (Fig. 2b), which exhibit, along with the nonlinear absorption of CdSe quantum wires, nonlinear changes in the transmission near the bulk CdSe absorption edge (about 700 nm).[3)] The 1.79-eV band in the differential transmission spectrum corresponds to transitions in bulk CdSe, and its blue shift and beaching at high pumping powers can be attributed to renormalization of the semiconductor band gap, filling of electron and hole energy bands, effects of the Coulomb screening, and filling of the exciton phase space.[14,15] Differential absorption spectra of pumped samples allow us to separate the nonlinear absorption in quantum wires due to ''saturation'' of excitonic transitions (see below). It seems that changes in the bulk absorption occur only in the band near the fundamental absorption edge after fast relaxation (in less than 10^{-11} s) of carriers to the band minima, whereas the discrete bleaching bands in the differential transmission spectra in the 1.6–2.2-eV band for GaAs and 1.9–2.1-eV band for CdSe are due to changes in the absorption of nanostructures. Thus, we can get rid of the background due to the bulk semiconductor by measuring differential transmission spectra.

In accordance with this theoretical model (see above), we attribute the features in the linear absorption spectra and bleaching bands in differential transmission spectra of samples with GaAs and CdSe wires at 2.14 eV, 1.84 eV, and 1.68 eV in Fig. 1, and 2.03 eV in Fig. 2 to linear and nonlinear absorption by $e-hh$ and $e-lh$ excitons (A and B excitons in CdSe) in quantum wires with diameters of 4.8 nm and 6.0 nm. Table 1 gives calculations of the exciton binding energy E_b and average length L_{exc} in quantum wires, positions of excitonic peaks in linear absorption spectra and, for comparison, their measurements, and also parameters of the structures used in our calculations, namely the quantum wire diameter d, effective masses of electrons and holes in the size quantization plane, $m_{\perp,e,h}$, the exciton reduced mass μ ($\mu^{-1}=m_{\|e}^{-1}+m_{\|h}^{-1}$), high-frequency permittivities[4)] of the semiconductor and dielectric, $\varepsilon_\infty^{w,b}$, the band gap E_g^0 of the bulk semiconductor and that renormalized due to the size quantization, E_g. We assume that the potential barriers for electrons and holes are equal and calculate them assuming a band gap of 4 eV in chrysotile-asbestos. For simplicity the effective masses in the semiconductor and dielectric are supposed to be equal, since their variation within reasonable limits changes E_g by 10–50 meV, which is smaller than line widths measured in the experiment. Note also that the average diameter of localized electrons and holes is much smaller than d, and the calculations yield approximately 2–3 nm.

One can clearly see in Table 1 that the size quantization and electrostatic self-action effect in such narrow quantum wires gives rise to a considerable renormalization of the semiconductor band gap, and the image potentials result in very large exciton binding energies. The latter are many times larger than the binding energies in the GaAs quantum wire formed on the crossing between two GaAs/GaAlAs quantum wells with a width of 7 nm (so-called T-shaped quantum wires, where $E_b=17$ meV).[3]

The calculations which are in good agreement with mea-

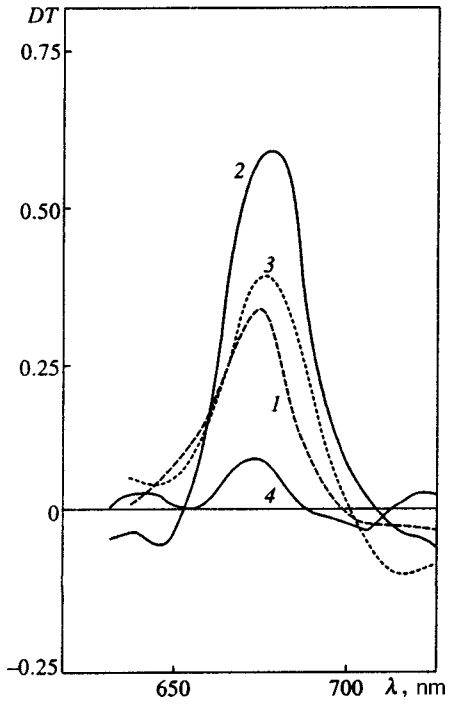


FIG. 3. Spectra of differential transmission of GaAs quantum wires (near the 1.84-eV absorption band) for different delays between the pumping and probing pulses: (*1*) 13 ps; (*2*) 0; (*3*) 20 ps; (*4*) 50 ps.

sured linear absorption spectra also allow us to analyze features of the nonlinear spectra. Hence the differential transmission spectrum 2 shown in Fig. 1 is typical of most regions in samples containing largely quantum wires 4.8 nm in diameter, whereas the spectrum 3 in Fig. 1 was obtained only in some fractions of the samples which contain, apparently, comparable numbers of both wire species. This fact is due to both the inhomogeneous distribution of nanotubes with diameters of 4.8 nm and 6.0 nm in the samples and the predominance of thinner wires in them. A similar picture was observed in samples with CdSe wires. In the latter samples, however, the energies of transitions due to *A* and *B* excitons are closer (the separation between them is within the spectral line width) owing to the smaller difference between masses of *A*- and *B*-hole subbands and initial separation between these subbands, as a result, these pairs of excitonic transitions form wider excitonic absorption bands (Fig. 2).

As the optical pumping intensity increases, different nonlinear processes in quantum wires co-exist and compete with one another, namely, the effect of filling of the phase space and screening of excitons, filling of electron and hole energy bands, and renormalization of the band gap in the one-dimensional semiconductor.[16] Generation of excitons and high-density plasma causes bleaching in the exciton absorption band (curves *2* and *3* in Fig. 1, and Figs. 2b and 3) because the electron–hole interaction is moderated owing to screening effects and filling of the phase space. The effect of phase space filling becomes apparent because excitons are composed of electrons and holes, which obey the Pauli exclusion principle, whereas only electron–hole states unoccupied by electrons and holes can be involved in forming excitons. The screening gives rise to a spatial redistribution of electrons and holes. This process, which is usually suppressed in quasi-one-dimensional structures, can be greatly intensified in the presence of image charges. Both these effects—filling of the phase space and Coulomb screening—shift the exciton band toward short wavelengths. On the other hand, the band gap renormalization taking place at high densities of photogenerated nonequilibrium carriers (it is up to $10^{-7}\,\mathrm{cm}^{-1}$ in our experiments) shift the exciton absorption line toward long wavelengths.

One can see in Fig. 3 that the 1.84-eV band in the differential transmission spectrum of the sample with GaAs is broadened, since its low-frequency edge shifts toward long wavelengths with the pumping power (spectra recorded at different delays between the pumping and probing pulses are compared). The long-wave shift depends on the delay and is maximum at zero delay (when the photogenerated plasma density is maximal). Thus, the band gap renormalization is the dominant effect in our samples.

Given the measurements of light-induced bleaching of samples with GaAs quantum wires crystallized in chrysolite-asbestos nanotubes in the 1.84-eV band, we can estimate the cubic component $\chi^{(3)}$ of the nonlinear susceptibility[5)] in our samples:[17]

$$\mathrm{Im}\,\chi^{(3)}=\frac{c^2 n_0^2 \Delta\alpha(\omega)}{8\pi^2\omega I(\omega)}, \tag{2}$$

where $\Delta\alpha$ is the change in the absorption coefficient in the sample subject to a resonant monochromatic field of intensity $I(\omega)$, and n_0 is the linear refraction index. With quasi-stationary pumping, $I(\omega)$ can be expressed in terms of the number n_e of photogenerated excitons in one quantum wire:

$$I(\omega)=\frac{n_e N l \hbar\omega}{[1-T_0(\omega)-r]\tau}, \tag{3}$$

where N is the density of quantum wires in the chrysotile-asbestos matrix, τ is the recombination time of photogenerated excitons, T_0 and r are the linear transmission and reflectivity of the sample, and l is its thickness. The spectrum of $\Delta\alpha$ can be derived from the spectrum $DT(\lambda)$:

$$\Delta\alpha=-\ln(1+DT)/l. \tag{4}$$

After substituting in Eq. (2) the maximum value of $\Delta\alpha$ at the excitonic transition frequency (in the 1.84-eV band) and the measured relaxation time of the light-induced change in the absorption, we obtain $\mathrm{Im}\,\chi^{(3)}\approx -4\times10^{-8}$ cgs units (the relaxation time is about 30 ps). This value of the dynamic nonlinearity is four to five orders of magnitude higher than the amplitude of classical noninertial nonlinearity. Note also that the relaxation time of this nonlinearity is quite short.

5. CONCLUSIONS

The behavior of linear and nonlinear absorption spectra of GaAs and CdSe semiconducting quantum wires imbedded in dielectric nanotubes has been interpreted in terms of excitonic transitions and their saturation (filling of the phase space and screening of excitons).

Owing to the effect of dielectric strengthening, the exciton binding energy has been observed to increase considerably (by more than 100 meV) in the samples studied. The theoretical estimates of transition energies in quantum wires provide a fairly accurate description of features in linear absorption spectra.

The work was supported by the Russian Fund for Fundamental Research (Grants 96-2-17339 and 97-2-17600), *Physics of Solid-State Nanostructures* program (Grants 97-1083 and 97-1072), and the *Materials for Electronic Techniques* project. We are indebted to V. V. Poborchii and S. G. Romanov for supplying the samples for our experiments, N. A. Kiselev and D. N. Zakharov for electronic microscope measurements, and R. Zimmerman for constructive criticism.

*)E-mail: tikh@gpi.ru

1)These measurements were performed by N. A. Kiselev and D. N. Zakharov (Institute of Crystallography, Russian Academy of Sciences).

2)The self-action potentials in quantum wires are derived from formulas for the electrostatic potential acting on charges within quantum wires (see Appendix to Ref. 8). Note that the factor $(-1)^n$ in Eqs. (A2) and (A5) should be omitted because it resulted from an error. Although this factor has no effect on the results of Ref. 8, in this specific case it is important.

3)The limited spectral range of the OVA-284 multichannel analyzer used in our experiments, however, has not allowed us to detect similar features in GaAs spectra (the optical absorption edge of bulk GaAs is around 860 nm).

4)We have used in our calculations high-frequency permittivities far from the phonon resonance, since the exciton energy is many times higher than the optical phonon energy.

5)We consider the so-called strong (resonant) dynamic (inertial) nonlinearity associated with light absorption in semiconducting nanostructures.[18] Unlike the so-called ''classical'' nonlinearity in transparent media, the strong nonlinearity occurs when high densities of quasiparticles (electrons, holes, excitons, etc.) are generated. The relaxation time of strong nonlinearity is determined by the lifetime of nonequilibrium quasiparticles. In 1926 S. I. Vavilov and V. L. Levshin detected strong optical nonlinearity in uranium glass subjected to resonant pumping[19] associated with saturation of the optical transition (saturation of a two-level system).

[1] Y. Arakawa and A. Yariv, IEEE J. Quantum Electron. **22**, 1887 (1986).
[2] W. Weigscheider, L. N. Pfeiffer, M. M. Dignam, A. Pinczuk, K. W. West, S. L. McCall, and R. Hull, Phys. Rev. Lett. **71**, 4071 (1993).
[3] S. Schmitt-Rink, D. A. B. Miller, and D. S. Chemla, Phys. Rev. B **35**, 8113 (1987).
[4] N. S. Rytova, Dokl. Akad. Nauk SSSR **163**, 1118 (1965) [Sov. Phys. Dokl. **10**, 754 (1966)].
[5] L. V. Keldysh JETP Lett. **29**, 658 (1979).
[6] M. S. Sundaram, S. A. Chalmers, P. F. Hopkins, and A. C. Gossard, Science **254**, 1326 (1991).
[7] V. V. Poborchii, M. S. Ivanova, and I. A. Salamatina, Superlattices Microstruct. **16**, 133 (1994); V. N. Bogomolov, Usp. Fiz. Nauk **124**, 171 (1978) [Sov. Phys. Usp. **21**, 77 (1978)]; N. V. Gushchina, V. S. Dneprovskii, E. A. Zhukov, O. V. Pavlov, V. V. Poborchii, and I. A. Salamatina, JETP Lett. **61**, 507 (1994); V. Dneprovskii, N. Gushina, O. Pavlov, V. Poborchii, I. Salamatina, and E. Zhukov, Phys. Lett. A **204**, 59 (1995).
[8] E. A. Muljarov and S. G. Tikhodeev, Zh. Éksp. Teor. Fiz. **111**, 274 (1997) [JETP **84**, 151 (1997)].
[9] E. A. Muljarov, S. G. Tikhodeev, N. A. Gippius, and T. Ishihara, Phys. Rev. B **51**, 14370 (1995).
[10] G. E. W. Bauer and T. Ando, Phys. Rev. B **38**, 6015 (1988).
[11] Landolt-Börnstein, *New Series, Group III*, Vol. 17, Springer-Verlag, Berlin (1982).
[12] T. Ogawa and T. Takagahara, Phys. Rev. B **43**, 14325 (1991); Phys. Rev. B **44**, 8138 (1991).
[13] S. Glutch and D. S. Chemla, Phys. Rev. B **53**, 15902 (1996).
[14] V. D. Egorov, Hoang Xuann Nguyen, R. Zimmermann, V. S. Dneprovskii, M. Kaschke, and D. S. Khechinashvili, Phys. Status Solidi B **159**, 403 (1990).
[15] H. Haug and S. W. Koch, *Quantum Theory and Electronic Properties of Semiconductors*, World Scientific, Singapore (1990).
[16] S. Benner and H. Haug, Europhys. Lett. **16**, 579 (1991).
[17] V. S. Dneprovskii, V. I. Klimov, D. K. Okorokov, and Yu. V. Vandyshev, Solid State Commun. **81**, 227 (1992).
[18] P. N. Butcher and D. Cotter, *The Elements of Nonlinear Optics*, Cambridge University Press (1990).
[19] S. I. Wawilow and W. L. Lewschin, Z. Phys. **35**, 932 (1926).

Translation provided by the Russian Editorial office.

Spatial–temporal dispersion of the kinetic coefficients near the Anderson transition

S. G. Novokshonov*) and A. G. Groshev

Physicotechnical Institute, Ural Branch of the Russian Academy of Sciences, 426001 Izhevsk, Russia
(Submitted 27 January 1998)
Zh. Éksp. Teor. Fiz. **114**, 711–724 (August 1998)

A generalization of the Vollhardt–Wölfle localization theory is proposed to make it possible to study the spatial–temporal dispersion of the kinetic coefficients of a d-dimensional disordered system in the low-frequency, long-wavelength range ($\omega \ll \mathscr{E}_F$ and $q \ll k_F$). It is shown that the critical behavior of the generalized diffusion coefficient $D(q,\omega)$ near the Anderson transition agrees with the general Berezinskiĭ–Gor'kov localization criterion. More precisely, on the metallic side of the transition the static diffusion coefficient $D(q,0)$ vanishes at a mobility threshold λ_c common for all q: $D(q,0) \propto t = (\lambda_c - \lambda)/\lambda_c \to 0$, where $\lambda = 1/(2\pi \mathscr{E}_F \tau)$ is a dimensionless coupling constant. On the insulator side, $q \neq 0$ $D(q,\omega) \propto -i\omega$ as $\omega \to 0$ for all finite q. Within these limits, the scale of the spatial dispersion of $D(q,\omega)$ decreases in proportion to t in the metallic phase and in proportion to $\omega\xi^2$, where ξ is the localization length, in the insulator phase until it reaches its lower limit $\sim \lambda_F$. The suppression of the spatial dispersion of $D(q,\omega)$ near the Anderson transition up to the atomic scale confirms the asymptotic validity of the Vollhardt–Wölfle approximation: $D(q,\omega) \simeq D(\omega)$ as $|t| \to 0$ and $\omega \to 0$. By contrast, the scale of the spatial dispersion of the electrical conductivity in the insulator phase is of order of the localization length and diverges in proportion to $|t|^{-\nu}$ as $|t| \to 0$. [S1063-7761(98)02608-0]

1. INTRODUCTION

The problem of the Anderson transition[1] is central to the theory of disordered systems (see the review in Refs. 2–6). The success achieved in this field of research is largely due to the development of self-consistent localization theory. Initially developed by Vollhardt and Wölfle[7,8] for low-dimensional systems ($d \leqslant 2$), it was later extended to systems of arbitrary dimension d (Refs. 9 and 10). The theory uses the well-developed apparatus of averaged Green's functions (Ref. 11, Sec. 39). The self-consistent approach of Vollhardt and Wölfle is based on an attractive idea about the nature of the localization phenomenon[12] and provides an effective interpolation method for calculating the kinetic coefficients of disordered systems, a method whose validity extends from classical kinetic theory to Anderson insulators. The main conclusions of the Vollhardt–Wölfle theory agree with the results of the field-theoretic[13] and scaling[3] approaches to the Anderson localization problem. What is also important is that this theory allows for generalizations that take into account various electron scattering mechanisms, the effect of external fields, and other factors (see, e.g., the review in Ref. 5).

All this makes self-consistent localization theory extremely useful from the practical viewpoint. However, it has a number of drawbacks, which often makes one question the reliability of its conclusions. A detailed analysis of the difficulties of the Vollhardt–Wölfle theory can be found in the review in Ref. 4 (see also Ref. 14). Some of these are discussed below, but here we touch on the problems that arise as a result of ignoring the spatial dispersion of the diffusion coefficient. The important point is that the basic equation of the Vollhardt–Wölfle theory establishes an integral relationship between the local diffusion coefficient $D(\omega) = D(q = 0,\omega)$ and the generalized diffusion coefficient $D(q,\omega)$. Usually this difficulty is avoided by assuming, as in Refs. 7 and 8, that $D(q,\omega) \simeq D(\omega)$. This, however, is an uncontrollable approximation.

The problem of the spatial dispersion of the kinetic coefficients near the Anderson transition has scarcely been studied at all to date.[5] Qualitative estimates of the q-dependence of $D(q,0)$ made on the basis of scaling considerations[3] lead to contradictory results and in fact destroy the structure of self-consistent localization theory. Incorporating into the σ-model Lagrangian terms with higher powers (>2) of gradients (which is equivalent to allowing for spatial dispersion of the diffusion coefficient) leads to an anomalous increase under scaling transformations,[15] i.e., to instability of the renormalization group. The approximations adopted in deriving the basic equations of the Vollhardt–Wölfle theory[7–9] do not make it possible to consistently take into account the spatial dispersion of the kinetic coefficients within the scope of the theory either.[5]

The first to examine this problem consistently was Suslov,[14] who concluded that in a small neighborhood of the Anderson transition the spatial dispersion of the diffusion coefficient becomes insignificant on scales $q \propto 1/\xi$ (ξ is the localization length), and its presence for $q \propto k_F$ (k_F is the Fermi momentum) does not influence the critical behavior of $D(q,\omega)$ predicted by self-consistent localization theory.[9,10] Actually this means that in the critical region ($|t| \to 0$ and $\omega \to 0$) the Vollhardt–Wölfle theory becomes asymptotically exact. The symmetry analysis of the Anderson transition done by Suslov[14] is extremely general, but its quantitative

results concerning the behavior of $D(q,\omega)$ in the critical region are valid only asymptotically, as $\omega\to 0$.

In a recent paper[16] we proposed a generalization of self-consistent localization theory that makes it possible to consistently take into account the spatial–temporal dispersion of the diffusion coefficient for the charge carriers in a two-dimensional disordered system at finite frequency and wave-number values in the region $\omega\ll\mathscr{E}_F$ and $q\ll k_F$ ($\mathscr{E}_F$ is the Fermi energy). At the microscopic level the results of that paper support the main conclusions drawn by Suslov,[14] who to a certain extent obtained them phenomenologically. In particular, there we show that due to suppression of the spatial dispersion of $D(q,\omega)$ the Vollhardt–Wölfle approximation $D(q,\omega)\simeq D(\omega)$ becomes valid in the localization regime but only in the region $\omega\tau\ll(l/\xi)^2(\lambda_F/l)^{1/3}\ll 1$ ($l=v_F\tau$ is the mean free path). Thus, meaningful calculations of $D(\omega)$ in a broad frequency range require consistent allowance for the q-dependence of the generalized diffusion coefficient. As a possible way of resolving this problem we propose a treatment of disordered systems with an arbitrary number of dimensions d that generalizes the approach developed in Ref. 16 for $d=2$.

2. STATEMENT OF THE PROBLEM AND THE BASIC EQUATIONS

We consider a d-dimensional degenerate ideal gas of spinless electrons subject to elastic scattering by immobile impurities having a concentration n_I and distributed in the sample according to the Poisson law. The one-electron Hamiltonian of the problem has the form

$$H=\frac{\mathbf{p}^2}{2m}+\sum_{\mathbf{R}} U(\mathbf{r}-\mathbf{R}), \tag{1}$$

where $U(\mathbf{r}-\mathbf{R})$ is the potential of an isolated impurity localized at the point with the radius vector $\mathbf{R}$. We assume that the potential is short-range and can be approximated by a delta function: $U(\mathbf{r})=U_0\delta(\mathbf{r})$. This is a good approximation provided that $r_0\ll\lambda_F, l$, where r_0 is the range of the potential $U(\mathbf{r})$, λ_F is the de Broglie wavelength, and l is the electron mean free path at the Fermi level. We also assume that the scattering of an electron by an isolated impurity is weak and that the first Born approximation is sufficient for calculating the scattering amplitude.

Since the system is spatially homogeneous on the average, the averaged one-electron Green's function is diagonal in the momentum representation:

$$\langle\langle\mathbf{p}|R^{\pm}(\mathscr{E})|\mathbf{p}'\rangle\rangle_I=\delta_{\mathbf{pp}'}G^{\pm}_{\mathbf{p}}(\mathscr{E})=\frac{\delta_{\mathbf{pp}'}}{\mathscr{E}-\mathscr{E}_{\mathbf{p}}-\Sigma^{\pm}_{\mathbf{p}}(\mathscr{E})}. \tag{2}$$

Here $R^{\pm}(\mathscr{E})=(\mathscr{E}-H\pm i\delta)^{-1}$ ($\delta\to+0$) is the resolvent of the Hamiltonian (1), $\langle\cdots\rangle_I$ indicates averaging over the impurity distribution, and $\Sigma^{\pm}_{\mathbf{p}}(\mathscr{E})$ is the electron self-energy part, which determines the perturbation of single-particle levels $\mathscr{E}_{\mathbf{p}}=\mathbf{p}^2/2m$ in the random field of the impurities.

The information on the kinetic properties of the system in the low-frequency and long-wavelength limits ($\omega\ll\mathscr{E}_F$ and $q\ll k_F$) is contained in the two-particle Green's function

$$\varphi_{\mathbf{pp}'}(\mathbf{q},\omega)=\langle\langle\mathbf{p}_+|R^+(\mathscr{E}^+)|\mathbf{p}'_+\rangle\langle\mathbf{p}'_-|R^-(\mathscr{E}^-)|\mathbf{p}_-\rangle\rangle_I \tag{3}$$

($\mathbf{p}_{\pm}=\mathbf{p}\pm\mathbf{q}/2$ and $\mathscr{E}^{\pm}=\mathscr{E}\pm\omega/2$), simply related to the correlation functions of the density–density (diffusion propagator) and the density–current type:

$$P(q,\omega)=\frac{1}{2\pi n_F}\sum_{\mathbf{pp}'}\varphi_{\mathbf{pp}'}(\mathbf{q},\omega)=\frac{1}{-i\omega+q^2D(q,\omega)}, \tag{4}$$

$$P_j(q,\omega)=\frac{1}{2\pi n_F}\sum_{\mathbf{pp}'}\frac{\hat{\mathbf{q}}\cdot\mathbf{p}}{m}\varphi_{\mathbf{pp}'}(\mathbf{q},\omega),\quad \hat{\mathbf{q}}=\frac{\mathbf{q}}{q},$$

where n_F is the density of states at the Fermi level, and $D(q,\omega)$ is the generalized diffusion coefficient. The correlation functions (4) satisfy the continuity equation

$$\omega P(q,\omega)-qP_j(q,\omega)=i+O\left(\frac{\omega}{\mathscr{E}_F}\right), \tag{5}$$

which is one way to mathematically formulate the law of conservation of the number of particles.

Using Eqs. (4) and (5), we can easily arrive at the following definition of the generalized diffusion coefficient:

$$D(q,\omega)=\frac{i}{q}\frac{P_j(q,\omega)}{P(q,\omega)}, \tag{6}$$

which is related to the electrical conductivity $\sigma(q,\omega)$ by the equation[17]

$$\sigma(q,\omega)=e^2n_F\frac{D(q,\omega)}{1-(q^2/i\omega)D(q,\omega)}. \tag{7}$$

The reason for the discrepancy between this equation and the ordinary Einstein relation $\sigma(\omega)=e^2n_FD(\omega)$ is that in the case of a spatially inhomogeneous nonequilibrium state the total longitudinal current, specified by the generalized Ohm's law $j(q,\omega)=\sigma(q,\omega)E(q,\omega)$, is the sum of the drift and diffusion currents: $j=j^{\text{drift}}+j^{\text{diff}}$. Indeed, in addition to the drift current, a longitudinal electric field $E(q,\omega)$ generates in the system a finite gradient of the charged-particle density, and hence a diffusion current $j^{\text{diff}}=(q^2/i\omega)D(q,\omega)j$.

We use the relation

$$\Delta G_{\mathbf{p}}(\mathbf{q},\omega)\Phi(\mathbf{p},\mathbf{q},\omega)=\sum_{\mathbf{p}'}\varphi_{\mathbf{pp}'}(\mathbf{q},\omega) \tag{8}$$

to determine a density relaxation function $\Phi(\mathbf{p},\mathbf{q},\omega)$ that satisfies the transport equation

$$\left[\omega-\frac{\mathbf{q}\cdot\mathbf{p}}{m}+\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)\right]\Phi(\mathbf{p},\mathbf{q},\omega)$$

$$-\sum_{\mathbf{p}'}U_{\mathbf{pp}'}(\mathbf{q},\omega)\Delta G_{\mathbf{p}'}(\mathbf{q},\omega)\Phi(\mathbf{p}',\mathbf{q},\omega)=1, \tag{9}$$

which can easily be obtained from the Bethe–Salpeter equation[7–9] for $\varphi_{\mathbf{pp}'}(\mathbf{q},\omega)$ of Eq. (3). Here we have introduced the notation

$$\Delta G_{\mathbf{p}}(\mathbf{q},\omega)=G^{-}_{\mathbf{p}_-}(\mathscr{E}^-)-G^{+}_{\mathbf{p}_+}(\mathscr{E}^+),$$
$$\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)=\Sigma^{-}_{\mathbf{p}_-}(\mathscr{E}^-)-\Sigma^{+}_{\mathbf{p}_+}(\mathscr{E}^+). \tag{10}$$

The self-energy part $\Sigma^{\pm}_{\mathbf{p}}(\mathscr{E})$ and the kernel $U_{\mathbf{pp}'}(\mathbf{q},\omega)$ (the irreducible vertex) of the integral equation (9) are related by the Ward identity[7–9]

$$\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)=\sum_{\mathbf{p}'} U_{\mathbf{pp}'}(\mathbf{q},\omega)\Delta G_{\mathbf{p}'}(\mathbf{q},\omega), \tag{11}$$

which plays an important role in the ensuing calculations; in particular, it guarantees that the transport equation (9) satisfies the law of conservation of the number of particles. Specifically, after multiplying Eq. (9) by $\Delta G_{\mathbf{p}}(\mathbf{q},\omega)$ we sum the product over $\mathbf{p}$, allowing for the Ward identity (11) and the symmetry property of the irreducible vertex, $U_{\mathbf{pp}'}(\mathbf{q},\omega)=U_{\mathbf{p}'\mathbf{p}}(\mathbf{q},\omega)$. As a result we arrive at the continuity equation (5), which relates the correlation functions (4).

In the self-consistent localization-theory approximation,[7–9] the irreducible vertex $U_{\mathbf{pp}'}(\mathbf{q},\omega)$ has the form

$$U_{\mathbf{pp}'}(\mathbf{q},\omega)=W+\frac{W}{\tau_0}\frac{1}{-i\omega+(\mathbf{p}+\mathbf{p}')^2D(|\mathbf{p}+\mathbf{p}'|,\omega)}, \tag{12}$$

where $W=n_IU_0^2$, $\tau_0=1/2\pi n_FW$ is the bare relaxation time (the lifetime) of an electron at the Fermi level, and $D(q,\omega)$ is the exact wave-number- and frequency-dependent diffusion coefficient. The diffusion pole in the second term on the right-hand side of Eq. (12) causes the elastic ($\omega\to0$) backscattering ($\mathbf{p}\to-\mathbf{p}'$) probability to diverge, which is the physical reason for the localization of charge carriers in disordered systems.[12] As shown by Suslov,[14] the singular structure of (12) is the result of the symmetry of the system under time reversal, and the exact expression for $U_{\mathbf{pp}'}(\mathbf{q},\omega)$ may differ from (12) only by factors in the first and second terms that are smooth functions of $\mathbf{p}$, $\mathbf{p}'$ and $\mathbf{q}$. However, allowing for these factors does not change the critical behavior of the diffusion coefficient near the Anderson transition,[14] and for this reason we use the approximation (12) in what follows.

At first glance the presence of a diffusion pole of type (12) in the irreducible vertex $U_{\mathbf{pp}'}(\mathbf{q},\omega)$ should lead to a divergence on the right-hand side of the Ward identity (11): $\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)\propto1/\omega$ in the insulator phase and $\Delta\Sigma_{\mathbf{p}}(\mathbf{q},0)\propto1/|t|$ in the static regime near the mobility threshold λ_c on the metallic side of the transition. This, however, contradicts the generally accepted viewpoint according to which the averaged single-particle Green's function, considered a function of the energy $\mathscr{E}$ and the frequency ω, retains its analytic properties in the metal–insulator transition.[2] This paradox was resolved in Ref. 14, where it was shown that in the right-hand side of the Ward identity (11) the divergences of type $1/\omega$ or $1/|t|$ cancel out because the singular part of $U_{\mathbf{pp}'}(q,\omega)$ is approximately (to within terms of order $O(\omega)$ or $O(|t|)$, respectively) orthogonal to $\Delta G_{\mathbf{p}}(\mathbf{q},\omega)$. The Vollhardt–Wölfle approximation (12) does not satisfy this condition and hence violates the Ward identity (11). The difficulties associated with this fact have been analyzed in a review by Sadovskiĭ.[4] Bearing all this in mind, we regard the Ward identity (11) in what follows as being a formal relationship between one- and two-particle characteristics of the system, and the electron self-energy part $\Sigma^{\pm}_{\mathbf{p}}(\mathscr{E})$ is assumed *a priori* to be an analytic function of its arguments. In this approach the lifetime $\tau=1/2\ \mathrm{Im}\,\Sigma^{-}_{\mathbf{p}}(\mathscr{E})$ must be considered a parameter of the theory, generally differing from the bare lifetime τ_0. However, following custom,[5,9,10] we will think of τ and τ_0 as being identical, since this does not change our results dramatically.

3. SOLVING THE TRANSPORT EQUATION

On the average, the system is isotropic, so that the relaxation function $\Phi(\mathbf{p},\mathbf{q},\omega)$ given by (8) depends on the absolute values of its vector arguments, $p=|\mathbf{p}|$ and $q=|\mathbf{q}|$, and the angle $\theta=\widehat{\mathbf{p},\mathbf{q}}$. It is therefore convenient to expand the function in a series in Gegenbauer polynomials (Ref. 18, Sec. 10.9):

$$\Phi(\mathbf{p},\mathbf{q},\omega)=\sum_{n=0}^{\infty}\frac{A_0}{A_n}\Phi_n(p,q,\omega)C_n^{(d-2)/2}(\cos\theta),$$
$$\Phi_n(p,q,\omega)=\frac{1}{A_0}\int_{-1}^{1}(1-x^2)^{(d-3)/2}C_n^{(d-2)/2}(x)\times\Phi(\mathbf{p},\mathbf{q},\omega)dx, \tag{13}$$

where $x=\cos\theta$, and A_n is a normalization constant. The system of Gegenbauer polynomials is a natural orthogonal basis for expanding functions that depend on the polar angle θ ($x=\cos\theta$) in a d-dimensional spherical system of coordinates. For $d=3$ the expression (13) coincides with the expansion in Legendre polynomials $P_n(x)$, and as $d\to2$ it becomes the expansion of the relaxation function in a Fourier series in $\cos n\theta$ (see Ref. 16).[1)]

In the definition (8), the delta-function singularity of the two-particle Green's function (3) at $p\simeq k_F$ ($q\ll k_F$) has been explicitly separated. Hence the expansion coefficients in (13) are functions that depend only weakly on p in the interval $|p-k_F|<1/l$. This makes it possible to approximately pass from the integral transport equation (9) to a system of linear algebraic equations with respect to $\Phi_n=\Phi_n(k_F,q,\omega)$. To this end we multiply (9) by $C_n^{(d-2)/2}(x)\Delta G_{\mathbf{p}}(\mathbf{q},\omega)/2\pi n_F$ and, after substituting the expansion (13) for $\Phi(\mathbf{p},\mathbf{q},\omega)$, sum over $\mathbf{p}$, allowing for the recurrence relations for the Gegenbauer polynomials. Using the delta-function property of $\Delta G_{\mathbf{p}}(\mathbf{q},\omega)$, we take all the functions that depend smoothly on p outside the summation signs. As a result we arrive at a system of equations for the expansion coefficients in (13) ($n\geqslant1$):

$$\omega\Phi_0-\frac{1}{d-2}\frac{qk_F}{m}\Phi_1=1,$$
$$\left(\omega+\frac{i}{\tau}\right)\Phi_n-\frac{qk_F}{m}\left[\frac{n+d-3}{2n+d-2}\Phi_{n-1}+\frac{n+1}{2n+d-2}\Phi_{n+1}\right]+\sum_{n'=1}^{\infty}M_{nn'}(q,\omega)\Phi_{n'}=0. \tag{14}$$

All coefficients in this system that are regular for $\omega\to0$ and $\mathbf{q}\to\mathbf{0}$ have been calculated to zeroth order in the smallness

parameters q/k_F and $\omega/\mathscr{E}_F$, e.g., $\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)\simeq\Delta\Sigma_{\mathbf{p}}(\mathbf{0},0)=i/\tau$. The elements of the memory function matrix $M_{nn'}(q,\omega)$ have the form

$$M_{nn'}(q,\omega)=-\frac{1}{2\pi i n_F}\frac{A_0}{A_{n'}}\sum_{\mathbf{pp'}} C_n^{(d-2)/2}(x)\Delta G_{\mathbf{p}}(\mathbf{q},\omega) \times U_{\mathbf{pp'}}(\mathbf{q},\omega)\Delta G_{\mathbf{p'}}(\mathbf{q},\omega)C_{n'}^{(d-2)/2}(x'), \tag{15}$$

where $x=\cos(\widehat{\mathbf{p},\mathbf{q}})$, and $x'=\cos(\widehat{\mathbf{p}',\mathbf{q}})$.

In deriving the system of equations (14) and (15) we have allowed for the fact that due to the Ward identity (11) the matrix elements $M_{0n}(q,\omega)$ and $M_{n0}(q,\omega)$ $(n\geqslant 0)$ and the corresponding matrix elements of $\Delta\Sigma_{\mathbf{p}}(\mathbf{q},\omega)$ cancel out. In particular, because of this the first equation in (14) is equivalent to the continuity equation (5). Accordingly, the first two coefficients in the sequence $\{\Phi_n(k_F,q,\omega)\}$ coincide, to within terms $O(q/k_F)$, with the correlation functions (4):

$$\Phi_0(k_F,q,\omega)\simeq -iP(q,\omega),$$
$$\Phi_1(k_F,q,\omega)\simeq -i\frac{m}{k_F}(d-2)P_j(q,\omega). \tag{16}$$

Taking into account the explicit form of the Gegenbauer polynomial $C_1^{(d-2)/2}(x)=(d-2)x$ and the fact that $A_0/A_1=d/(d-2)^2$, we can easily prove that $M_{11}(q,\omega)$ of (15) coincides with the known expression for the current relaxation kernel.[7–9] Hence, keeping only the first two terms in the expansion (13) $(n=0,1)$ and allowing for (16), we see that (14) and (15) reduce to the closed system of equations of the Vollhardt–Wölfle theory.[7–9] As shown in Ref. 16, such an approximation makes it possible to calculate the kinetic coefficients without allowing for their spatial dispersion. Otherwise we are forced to solve the infinite system of equations (14). This can be done in the long-wavelength limit, since for $q\ll k_F$ the approximation linear in q/k_F is sufficient for calculating the elements of the memory function matrix (15), which in this approximation is tridiagonal:

$$M_{nn'}(q,\omega)=-\frac{(-1)^n}{\tau}\left\{i\Delta_n(\omega)\delta_{nn'} + ql\left[\frac{n+d-3}{2n+d-2}\Lambda_{n-1}(\omega)\delta_{nn'+1} - \frac{n+1}{2n+d-2}\Lambda_n(\omega)\delta_{nn'-1}\right]\right\}. \tag{17}$$

In the Appendix the coefficients $\Delta_n(\omega)$ and $\Lambda_n(\omega)$ are calculated approximately in the low-frequency $\omega\ll\mathscr{E}_F$ and long-wavelength $(q\ll k_F)$ regions.

In contrast to the regular coefficients of the system of equations (14), calculated in the zeroth order in the small parameters $\omega/\mathscr{E}_F$ and q/k_F, here we allow for the off-diagonal elements $M_{n,n\pm1}\propto q/k_F$. The point is that as $\omega\to 0$ in the insulator phase or as $|t|\to 0$ on the metallic side of the Anderson transition, in the static regime the singular parts of $M_{n,n\pm1}$ increase and in the corresponding limits begin to prevail over the terms in the square brackets of the second equation in (14). The singularities die to the diffusion pole in (12) are also present in the other elements of the memory function matrix. Nevertheless, in the long-wavelength limit the tridiagonal part (17) plays the leading role in the matrix. Indeed, if we continue to expand (15) in powers of q/k_F, the following estimate holds:

$$\left|\frac{M_{n,n\pm k}(q,\omega)}{M_{n,n\pm1}(q,\omega)}\right|\propto\left(\frac{q}{k_F}\right)^{k-1}\ll 1. \tag{18}$$

After (17) is substituted in (14), the infinite system of equations for the coefficients Φ_n becomes tridiagonal. Its formal exact solution can be obtained if we employ the methods of continued-fraction theory.[19] Actually it is enough to find the coefficient ratio Φ_1/Φ_0, in terms of which we can express, via Eqs. (6) and (16), the generalized diffusion coefficient. To this end, introducing the notation $Y_n=\Phi_{n+1}/\Phi_n$, we write the second equation in (14) as follows:

$$Y_{n-1}=-\frac{iql[1-(-1)^{n-1}\Lambda_{n-1}(\omega)](n+d-3)/(2n+d-2)}{1-i\omega\tau-(-1)^n\Delta_n(\omega)+iql[1-(-1)^n\Lambda_n(\omega)]Y_n(n+1)/(2n+d-2)}. \tag{19}$$

This begins the recurrence process, which makes it possible to write $D(q,\omega)\propto Y_0$ in the form

$$D(q,\omega)=\frac{D_0}{1-i\omega\tau+\Delta_1(\omega)}K(q,\omega), \tag{20}$$

where $D_0=v_F^2\tau/d$ is the classical diffusion coefficient of a d-dimensional system. The spatial dispersion of the generalized diffusion coefficient $D(q,\omega)$ of Eq. (20) is completely determined by the continued-fraction expansion

$$K(q,\omega)=\cfrac{1}{1+\cfrac{R_1^2(\omega)q^2}{1+\cfrac{R_2^2(\omega)q^2}{1+\cdots}}},$$

$$R_n^2(\omega)=\frac{(n+1)(n+d-2)}{(2n+d-2)(2n+d)}\times\frac{l^2[1-(-1)^n\Lambda_n(\omega)]^2}{[1-i\omega\tau-(-1)^n\Delta_n(\omega)][1-i\omega\tau-(-1)^{n+1}\Delta_{n+1}(\omega)]}. \tag{21}$$

Note that at $d=1$ we have $K(q,\omega)\equiv 1$, i.e., in the present approximation the q-dependence of the diffusion coefficient of a one-dimensional disordered system is important only on the atomic scale ($q\lambda_F\simeq 1$).[2)] According to this remark, the nonlocal nature of the diffusion coefficient on larger scales ($q\lambda_F\ll 1$) may manifest itself only in systems with $d>1$. In this case the range where Eqs. (20) and (21) are valid is limited by the convergence condition for the continued-fraction expansion of $K(q,\omega)$. According to the Worpitski test for convergence (Ref. 19, p. 107), this expansion tends to a finite limit if $q|R_n(\omega)|\leqslant 1/2$ for all $n>n_0\geqslant 1$. Since the singular coefficients $|\Delta_n(\omega)|\to 0$ and $|\Lambda_n(\omega)|\to 0$ as $n\to\infty$, the above condition is equivalent to the inequality $ql/|1-i\omega\tau|<1$.

4. CRITICAL BEHAVIOR NEAR THE ANDERSON TRANSITION

The continued-fraction expansion in (20) and (21), whose coefficients are determined by (A1) and (A3), constitutes a nonlinear integral equation with respect to $D(q,\omega)$. Replacing (A4) with (A3) yields the first iteration of the generalized diffusion coefficient, with $\tilde{D}(\omega)=D((i\omega/\tilde{D})^{1/2},\omega)\simeq 1$ as the initial condition. The auxiliary quantity satisfies the self-consistency equation obtained from (20) if in the right-hand side we put $q^2=i\omega/\tilde{D}$. In the critical region we have ($\omega\to 0,|t|\to 0$) $\tilde{K}(\omega)=K((i\omega/\tilde{D})^{1/2},\omega)\simeq 1$, so that

$$\tilde{D}(\omega)=\frac{D_0}{1+\Delta_1(\omega)}. \tag{22}$$

Thus, the behavior of $\tilde{D}(\omega)$ (and hence of $D(q,\omega)$) near the Anderson transition is determined by a single parameter, $\Delta_1(\omega)$ of (A7), proportional to the current relaxation kernel.[7–9] For our purposes it is sufficient to use the low-frequency asymptotic expression

$$\Delta_1(\omega)=\frac{D_0}{\tilde{D}(\omega)}\left[f_d(\lambda)+\frac{2d\lambda^{d-1}}{d-2}(K_d y-C_d y^{(d-2)/2})\right] \tag{23}$$

($|y|\ll 1$), where

$$y=-4di\omega\tau\frac{D_0}{\tilde{D}(\omega)},\quad C_d=\left(\frac{\pi}{2}\right)^{d-2}\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{4-d}{2}\right),$$

$$K_d=\frac{C_d}{\sqrt{\pi}}\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{5-d}{2}\right).$$

Generally, to calculate the first term in (23) we must integrate (A1) with (A4) at $\omega=0$. Only in systems with d close to 2 will the mobility threshold, found from the equation $f_d(\lambda)=1$, satisfy the weak-coupling condition ($\lambda_c\ll 1$). In this range of values of λ and d we have a simple approximation for $f_d(\lambda)$:

$$f_d(\lambda)=\frac{4}{\pi}\frac{d}{d-2}\left(\frac{\pi}{2}\lambda\right)^{d-1},\quad \lambda\ll 1,\quad d\to 2+. \tag{24}$$

The mobility threshold calculated via (22) and (24),

$$\lambda_c=\frac{2}{\pi}\left(\frac{\pi}{4}\frac{d-2}{4}\right)^{1/(d-1)},\quad d\to 2+, \tag{25}$$

differs from the Vollhardt–Wölfle result (see Eq. (38) in Ref. 9) by an insignificant factor, $2^{(d-3)/(d-1)}$. The dependence of λ_c on d, which can be found by numerically solving the equation $f_d(\lambda_c)=1$, agrees quantitatively with (25) over the entire range $2<d<4$ (the relative deviation amounts to no more than 15–20%) and asymptotically tends to (25) as $d\to 2$. This agreement can be considered satisfactory. To analyze the critical behavior of the kinetic coefficients it is enough to expand $f_d(\lambda)$ in a Taylor series in the neighborhood of λ_c and keep only the terms linear in $t=(\lambda_c-\lambda)/\lambda_c$. This leads to the well-known asymptotic behavior[6,9,10]

$$\frac{\tilde{D}(\omega)}{D_0}\propto\begin{cases}(-i\omega\tau)^{1/(1+2\nu)}, & \omega\gg\omega_c\ \text{(metal–insulator)},\\ t, & \omega\ll\omega_c,\ t>0\ \text{(metal)},\\ -i\omega\xi^2, & \omega\ll\omega_c,\ t<0\ \text{(insulator)},\end{cases} \tag{26}$$

where $\omega_c\tau=|t|^{1+2\nu}$, and ν is the critical index of the localization length $\xi\propto|t|^{-\nu}$; $\nu=1/(d-2)$ for $2<d<4$, and $\nu=1/2$ for $d>4$. A detailed analysis of the critical behavior of the diffusion coefficient $D(\omega)$ can be found in Ref. 9 (see also the reviews in Refs. 4–6).

Now let us discuss the spatial dispersion of the kinetic coefficients in the critical region. To this end we approximate the continued fraction $K(q,\omega)$ in (21) by truncating it at the first denominator. As a result we arrive at an expression for the diffusion coefficient:

$$D(q,\omega)=\frac{\tilde{D}(\omega)}{1+R_1^2(\omega)q^2}, \tag{27}$$

which is valid if $|R_1(\omega)q|\ll 1$, where $R_1(\omega)$ is the radius of nonlocality [determining the scale of spatial dispersion of $D(q,\omega)$] defined in (21). The critical behavior of the generalized diffusion coefficient (27) is determined by the q-independent parameter $\tilde{D}(\omega)$ of (26). Hence $\lim_{\omega\to 0}D(q,\omega)=0$ in the insulator phase and $\lim_{t\to 0}D(q,0)=0$ in the metallic phase hold simultaneously for all values of q. Thus result, first obtained by Suslov[14] on the basis of a sym-

metry approach to the Anderson-transition problem, agrees with the Berezinskiĭ–Gor'kov localization criterion.[20]

How does the nonlocality range $R_1(\omega)$ of the generalized diffusion coefficient change when we replace the classical conductor by an Anderson insulator? Far from the Anderson transition we have $|\Delta_1(\omega)| \simeq |\Delta_2(\omega)| \ll 1$ and $|\Lambda_1(\omega)| \ll 1$ in the metallic phase, so that $R_1(\omega) \propto l \gg \lambda_F$ is of order of the mean free path or the diffusion length $l_D = \sqrt{D_0 \tau} = l/\sqrt{d}$.

The behavior of the nonlocality range of $D(q,\omega)$ near the Anderson transition depends on the relationship between the singular parameters $\Delta_n(\omega)$ and $\Lambda_n(\omega)$. In the critical region ($|t| \to 0$ and $\omega \to 0$) we have $\Delta_1(\omega) \simeq \Delta_2(\omega) \simeq D_0/\tilde{D}(\omega)$ ($|D_0/\tilde{D}(\omega)| \gg 1$), which causes suppression of the spatial dispersion of $D(q,\omega)$ as long as $|R_1(\omega)| \gg \lambda_F$ and $|\Lambda_1(\omega)| \ll 1$ hold simultaneously. Directly from (21) and the estimate (see (A8)) $|\Lambda_1(\omega)| \propto \lambda^k |D_0/\tilde{D}(\omega)|$ ($k=3$ at $d=3$, and $k=4$ at $d=2$ and $d=4$) it follows that both inequalities hold only when the disorder is extremely weak, when $\lambda \ll |\tilde{D}(\omega)/D_0| \ll 1$. In this case the asymptotic expression, obtained in Ref. 16 for a two-dimensional disordered system,

$$R_1^2(\omega) = -\frac{2(d-1)}{d(d+2)} l^2 \left(\frac{\tilde{D}(\omega)}{D_0} \right)^2, \tag{28}$$

is valid. According to (28) and (26), in the scaling regime ($\omega \gg \omega_c$) the nonlocality of the generalized diffusion coefficient is the same in the metallic and insulator phases. As we move into the critical region ($\omega \ll \omega_c$, $|t| \to 0$), together with suppression of the spatial dispersion of $D(q,\omega)$ there is a dramatic change in nature of the dispersion in the metallic phase ($R_1^2(\omega) < 0$) as compared to the insulator phase ($R_1^2(\omega) > 0$). However, in all cases the scale of the q-dependence of the generalized diffusion coefficient (28) is determined, in accordance with the physical interpretation proposed in Ref. 16, by the renormalized diffusion length $l_D(\omega) \propto l|\tilde{D}(\omega)/D_0|$.

In a three-dimensional system, the coupling constant at the mobility threshold (Eq. (25)) is equal to $\lambda_c \approx 0.32$, which corresponds to $l \approx 0.16\lambda_F$. With such a degree of disorder, the nonlocality range $R_1(\omega)$ given by Eqs. (28) and (26) is sure to be smaller, in absolute value, than λ_F. As noted at the end of Sec. 3, under these conditions the scale of the spatial dispersion of the generalized diffusion coefficient reaches its lower limit $|R_1(\omega)| \propto \lambda_F$, determined by the wave nature of the quantum mechanical laws of motion. In other words, the Anderson transition lies outside the limits of validity of the asymptotic expressions (28) and (26). Nevertheless, the anomalies in the spatial dispersion of $D(q,\omega)$ predicted by these expression may be observed in highly anisotropic (quasi-two-dimensional) disordered systems, in which the mobility threshold lies within the weak-disorder range $\lambda_c \ll 1$ (see Ref. 21).

In contrast to $D(q,\omega)$, the electrical conductivity $\sigma(q,\omega)$ specified by Eq. (7) and the related longitudinal dielectric constant $\varepsilon(q,\omega) = 1 + 4\pi i \sigma(q,\omega)/\omega$ exhibit much stronger anomalies in spatial dispersion, which are due to the presence of a diffusion pole in (7), Indeed, if we plug (27) into (7), we immediately obtain $|R_1(\omega)|^2 \ll |\tilde{D}(\omega)/i\omega|$ near the Anderson transition and hence

$$\varepsilon(q,\omega) = 1 + \frac{q_{FT}^2 \tilde{D}(\omega)}{-i\omega + q^2 \tilde{D}(\omega)}, \tag{29}$$

where $q_{FT}^{-1} = (4\pi e^2 n_F)^{-1/2}$ is the Fermi–Thomas screening radius.

Equation (29) is usually employed as the initial expression[22] in analyzing the dielectric and optical properties of disordered systems. The above results show that irrespective of the relationship between ω and $q^2|\tilde{D}(\omega)|$ it is valid because of suppression of the spatial dispersion of the generalized diffusion coefficient in Anderson localization. Substituting (26) in (29) yields the well-known asymptotic expressions for the dielectric constant.[4,22] In particular, in the critical region ($\omega \ll \omega_c$ and $|t| \to 0$),

$$\varepsilon(q,\omega) = 1 + \frac{q_{FT}^2 (\xi^2 + iD^{\text{hop}}/\omega)}{1 + (\xi^2 + iD^{\text{hop}}/\omega) q^2}, \tag{30}$$

holds on the insulator side of the transition, where D^{hop} is the hopping diffusion coefficient. As $\omega \to 0$, the above expression (30) becomes $\varepsilon(q,0) = 1 + q_{FT}^2/q^2$, i.e., in an Anderson insulator a static electric field is screened, as it is in an ordinary metal.[4]

At finite frequencies $\omega\xi^2 \gg |D^{\text{hop}}|$ holds in a small neighborhood of the Anderson transition, so that the contribution of hopping transfer to (30) can be ignored. In this case, shielding of the electric fields occurs only at distances which are small compared with the localization length, i.e., for $q\xi \gg 1$ (which is equivalent to $\omega \ll q^2|\tilde{D}(\omega)|$). In the opposite limit, i.e., $q\xi \ll 1$ ($\omega \gg q^2|\tilde{D}(\omega)|$), (30) tends to $\varepsilon(0,0) = 1 + q_{FT}^2 \xi^2$, the dc dielectric constant of a gas of neutral atoms with a concentration $e^2 n_F/\xi$ and a polarizability ξ^3. In other words, a long-wavelength low-frequency ($q\xi \ll 1$ and $\omega \ll \omega_c$) external field E^{ext} induces in an Anderson insulator a field $E = E^{\text{ext}}/\varepsilon(0,0)$, which decreases in proportion to $1/\varepsilon(0,0) \propto |t|^{2\nu}$ as the mobility threshold is approached until it reaches its minimum value $\sim (q/q_{FT})^2 E^{\text{ext}}$. The huge values of $\varepsilon(0,0)$ associated with this anomaly have been observed on the insulator side of the metal–insulator transition in, e.g., Si:P (Ref. 23).

5. CONCLUSION

Thus, within the scope of self-consistent localization theory, it has proved possible to consistently take into account the spatial dispersion of kinetic coefficients. This is important because by completely discarding the nonlocal behavior of the diffusion coefficient in the self-consistency equation (20) we end up with an uncontrollable approximation. Hence, strictly speaking, the Vollhardt–Wölfle theory[7–9] is valid only asymptotically, as $\omega \to 0$ (see Ref. 14). Our results make possible a quantitative analysis of the frequency dependence of the kinetic coefficients near the Anderson transition at finite frequencies satisfying the con-

dition $\omega \ll \mathscr{E}_F$. This problem arises, for instance, when the electron–electron interaction is taken into account in self-consistent localization theory.[5,24,25]

The nature of the spatial dispersion of $D(q,\omega)$ in Anderson localization strongly depends on the number of dimensions d of the disordered system. For $d=3$ the mobility threshold is in the strong-coupling region ($\lambda_c \approx 0.32$) and the nonlocality range $R_1(\omega)$ given by (28) is smaller than λ_F. In other words, near the Anderson transition, within a broad frequency range, the nonlocality of the diffusion coefficient of a three-dimensional system is important only on the atomic scale. Hence, ignoring the spatial dispersion in the self-consistency equation[25] is justified. The situation is different when we are dealing with low-dimensional systems, where the localization conditions are met in the weak-coupling range ($\lambda \ll 1$). As shown in Ref. 16, when $d=2$, we can ignore the spatial dispersion of $D(q,\omega)$ in the self-consistency equation only in the low-frequency limit $\omega\tau \ll (l/\xi)^2 \lambda^{1/3} \ll 1$, since otherwise, even when calculating $D(q=0,\omega)=D(\omega)$, we must use numerical methods to solve an integral (in $\mathbf{q}$) equation of type (20).

The authors are grateful to A. K. Arzhnikov for stimulating discussions and N. V. Sadovskiĭ and É. Z. Kuchinskiĭ for useful comments made in the discussion of the results of the present study.

APPENDIX: CALCULATION OF ELEMENTS OF THE MEMORY FUNCTION MATRIX

The contribution of the first term in (12) to the matrix element $M_{nn'}(q,\omega)$ (Eq. (15)) is nonsingular and of order $((q/k_F)^{n+n'})$ $(n,n' \geqslant 1)$, so it can be ignored. To calculate the singular part of the memory function matrix, we expand the second term in (12) in a series of the form (13) in Gegenbauer polynomials dependent on $\cos\gamma$, where $\gamma = \widehat{\mathbf{p},\mathbf{p}'}$ is the scattering angle. In the d-dimensional spherical system of coordinates we direct the polar axis parallel to the vector $\mathbf{q}$. Then $\cos\gamma = \cos\theta\cos\theta' + \cos\rho\sin\theta\sin\theta'$, where ρ is the angle between the planes containing the vector pairs $\mathbf{p},\mathbf{q}$ and $\mathbf{p}',\mathbf{q}$, respectively, and $\theta = \widehat{\mathbf{p},\mathbf{q}}$ and $\theta' = \widehat{\mathbf{p},\mathbf{q}'}$. Using the addition theorem (Ref. 26, Sec. 16.3, (20)) and carrying out simple transformations, we obtain

$$M_{nn'}(q,\omega) = -\frac{i}{\pi(d-2)\tau_0^2} \sum_{l=0}^{\infty} \frac{l!}{\Gamma(l+d-2)} \frac{1}{(2\pi)^d} \times \int_0^\infty p^{d-1}dp \int_0^\infty p'^{d-1}dp' P_l(p,p') \times \langle n|\Delta G_{\mathbf{p}}(\mathbf{q},\omega)|l\rangle \langle l|\Delta G_{\mathbf{p}'}(\mathbf{q},\omega)|n'\rangle, \quad \text{(A1)}$$

where the matrix elements of the difference of single-particle Green's functions are

$$\langle n|\Delta G_{\mathbf{p}}(\mathbf{q},\omega)|l\rangle = \frac{1}{2\pi i n_F A_l} \int_{-1}^{1} (1-x^2)^{(d-3)/2} C_n^{(d-2)/2}(x) \times C_l^{(d-2)/2}(x) \Delta G_{\mathbf{p}}(\mathbf{q},\omega)dx, \quad \text{(A2)}$$

and the coefficients in the expansion (13) of the diffusion propagator are

$$P_n(p,p') = \int_{-1}^{1} \frac{(1-x^2)^{(d-3)/2} C_n^{(d-2)/2}(x)dx}{-i\omega + (\mathbf{p}+\mathbf{p}')^2 D(|\mathbf{p}+\mathbf{p}'|,\omega)}, \quad \text{(A3)}$$

with $x = \cos\gamma$.

In a space with $d \leqslant 3$, the integrand in (A3) has a nonintegrable singularity as $\omega \to 0$ and $\mathbf{p} \to -\mathbf{p}'$. Hence, assuming that in the low-frequency limit the main contribution to (A3) is provided by the vicinity of the diffusion pole, we use the approximation in which the diffusion coefficient $D(|\mathbf{p}+\mathbf{p}'|,\omega)$ is replaced by its value at the pole, $\tilde{D}(\omega) = D((i\omega/\tilde{D})^{1/2},\omega)$ (see Ref. 16), after which we can use a tabulated integral (Ref. 26, Sec. 16.3, (17)) to express $P_n(p,p')$ in terms of the Legendre function of the second kind (Ref. 27, Chap. 3):

$$P_n(p,p') \simeq \frac{(-1)^n}{2pp'\tilde{D}(\omega)} \frac{2\sqrt{\pi}}{\Gamma((d-2)/2)} \exp\left\{-i\pi\frac{d-3}{2}\right\} \times \left(\frac{z^2-1}{4}\right)^{(d-3)/4} Q_{n+(d-3)/2}^{(d-3)/2}(z), \quad \text{(A4)}$$

where the Legendre function depends on the parameter $z = (-i\omega + (p^2+p'^2)\tilde{D}(\omega)) \times [2pp'\tilde{D}(\omega)]^{-1}$. Thus, formula (A4) determines the low-frequency asymptotic behavior of $P_n(p,p')$ when $d \leqslant 3$. For instance, for a two-dimensional system this formula is valid for $|\tilde{D}(\omega)/D_0|^3 \ll \lambda_F/l$ (Ref. 16). In systems with $d>3$, the singularity of the diffusion propagator in (A3) is balanced by the factor $(1-x^2)^{(d-3)/2}$, which originates in the Jacobian of the d-dimensional spherical system of coordinates. The integral in (A3) remains convergent as $z \to 1$; nevertheless, at low frequencies $\omega \ll k_F^2|\tilde{D}(\omega)|$ the leading contribution to it is still provided by a small neighborhood of the diffusion pole. As $\omega \to 0$, in the metallic phase ($\tilde{D}(0) \neq 0$) this condition is met automatically, while in the insulator phase ($\tilde{D}(\omega) \propto -i\omega\xi^2$) it is met if $k_F\xi \gg 1$.

The leading contribution to the integrals in (A1) comes from the vicinity of the Fermi level, $|p-k_F| \leqslant 1/l$, inside which $|z| \simeq 1$ if $\omega \ll k_F^2|\tilde{D}(\omega)|$. In this case, expressing $Q_\nu^\mu(z)$ in terms of the hypergeometric function (see Ref. 27, Sec. 3.2, (38)), we readily arrive at an expression for the asymptotic behavior of (A4):

$$P_n(p,p') \simeq \frac{(-1)^n}{2pp'\tilde{D}(\omega)} \frac{2\sqrt{\pi}\Gamma((5-d)/2)}{\Gamma((d-2)/2)} \frac{1}{d-3} \times \left[\frac{\Gamma((d-2)/2)}{\Gamma((5-d)/2)} - \frac{\Gamma(n+d-2)}{n!} \times \left(\frac{z^2-1}{4}\right)^{(d-3)/2}\right], \quad \text{(A5)}$$

which is valid in the neighborhood $|n^2(1-z^2)| \ll 1$ of the point $z=1$. This inequality means that (A5) can be used to obtain asymptotic estimates of the coefficients $M_{nn'}(q,\omega)$

only when $n<k_F l$. For larger values of n the asymptotic expression breaks down already inside the integration interval $|p-k_F|\simeq 1/l$ and (A1) can be calculated only by numerical methods.

Expanding the matrix elements (A2) of the single-particle Green's functions in a power series in $\mathbf{p}\cdot\mathbf{q}/m$ and limiting ourselves to first order in q,

$$\langle n|\Delta G_{\mathbf{p}}|l\rangle=\frac{1}{2\pi i n_F}\left\{\delta_{n,l}\Delta G-\frac{qp}{2m}\left[\frac{n+d-3}{2n+d-2}\,\delta_{n,l+1}+\frac{n+1}{2n+d-2}\,\delta_{n,l-1}\right]\frac{\partial\Delta G}{\partial\alpha}\right\}, \tag{A6}$$

where $\Delta G\simeq 2\alpha/[(\mathscr{E}-\mathscr{E}_{\mathbf{p}})^2+\alpha^2]$, $\alpha=\pi\mathscr{E}_F\lambda$, we arrive at a long-wavelength asymptotic expression for the tridiagonal memory function matrix (17). If we take into account the delta-function property of the diagonal part of (A6), one of the integrals in (A1) can be dropped. Substituting (A4) and (A5), we can evaluate the remaining integral analytically.

After some lengthy transformations, we can write the coefficient $\Delta_n(\omega)$ in (17) as follows ($\lambda\ll 1$):

$$\Delta_n(\omega)=\pi d\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{4-d}{2}\right)\lambda^2\,\frac{D_0}{\tilde{D}(\omega)}\,\frac{1}{3-d}\times\left[\left(\frac{\pi}{2}\lambda\right)^{d-3}F\left(\frac{3-d}{2},\frac{4-d}{2};\frac{5-d}{2};1-y\right)-\frac{n!\Gamma(d-2)}{\Gamma(n+d-2)}\sin\left(\pi\,\frac{d-2}{2}\right)\right], \tag{A7}$$

where $y=-4di\omega\tau D_0/\tilde{D}$, and $F(a,b;c;z)$ is Gauss's hypergeometric function (see Ref. 27, Chap. 2). At $d=2$ the singular part of (A7) coincides with the expression for Δ in Ref. 16. Reasoning along similar lines, we arrive at

$$\Lambda_n(\omega)=\frac{\pi^2}{2}\,d\lambda^3\,\frac{D_0}{\tilde{D}(\omega)}\,\frac{n!\Gamma(d-1)}{\Gamma(n+d-1)}\times\Gamma\left(\frac{d-1}{2}\right)\Gamma\left(\frac{5-d}{2}\right)\cos\left(\frac{d-1}{2}\,t\right), \tag{A8}$$

which is also valid for $\lambda\ll 1$, with $\cos t=-(1+\pi^2\lambda^2)^{-1/2}$.

Using the formulas for the analytic continuation of the hypergeometric function (Ref. 27, Sec. 2.10, (1)), we can easily derive the asymptotic expressions (23) and (24) from (A7).

*)E-mail: nov@otf.fti.udmurtia.su

1)From now on the number of dimensions d of the space can be considered a continuously varying real parameter. Its range is limited by the inequality $d>1$, which follows from the condition of the integrability of the weighting function $w(x)=(1-x^2)^{(d-3)/2}$ of the system of Gegenbauer polynomials (see Ref. 18, Sec. 10.9).

2)Due to the wave nature of the quantum mechanical laws of motion, λ_F is the lower bound on the nonlocality range, which determines the spatial dispersion of the kinetic coefficients.

[1] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[2] A. L. Éfros, Usp. Fiz. Nauk **126**, 41 (1978) [Sov. Phys. Usp. **21**, 746 (1978)].
[3] P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
[4] M. V. Sadovskiĭ, in *Soviet Scientific Reviews–Physics Reviews*, Vol. 7, I. M. Khalatnikov (ed.), Harwood Academic, New York (1985), p. 1.
[5] M. V. Sadovskiĭ, Sverkhprovodimost': Fiz. Khim. Tekhnol. **8**, 337 (1995).
[6] D. Vollhardt and P. Wölfle, in *Electronic Phase Transitions*, W. Hanke and Yu. V. Kopaev (eds.), North-Holland, Amsterdam (1992), p. 1.
[7] D. Vollhardt and P. Wölfle, Phys. Rev. Lett. **45**, 842, 1370 (1980).
[8] D. Vollhardt and P. Wölfle, Phys. Rev. B **22**, 4666 (1980).
[9] D. Vollhardt and P. Wölfle, in *Anderson Localization*, Y. Nagaoka and H. Fukuyama (eds.), Springer, Berlin (1982), p. 26.
[10] A. V. Myasnikov and M. V. Sadovskiĭ, Fiz. Tverd. Tela (Leningrad) **24**, 3569 (1982) [Sov. Phys. Solid State **24**, 2033 (1982)].
[11] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinski, *Quantum Field Theoretical Methods in Statistical Physics*, Pergamon Press, New York (1965).
[12] L. P. Gor'kov, A. I. Larkin, and D. E. Khmel'nitskiĭ, JETP Lett. **30**, 228 (1979).
[13] F. J. Wegner, in *Anderson Localization*, Y. Nagaoka and H. Fukuyama (eds.), Springer, Berlin (1982), p. 8.
[14] I. M. Suslov, Zh. Éksp. Teor. Fiz. **108**, 1686 (1995) [JETP **81**, 925 (1995)].
[15] V. E. Kravtsov, I. V. Lerner, and V. I. Yudson, Zh. Éksp. Teor. Fiz. **94**, No. 7, 255 (1988) [Sov. Phys. JETP **67**, 1441 (1988)].
[16] A. G. Groshev and S. G. Novokshenov, Zh. Éksp. Teor. Fiz. **111**, 1787 (1997) [JETP **84**, 978 (1997)].
[17] D. N. Zubarev, ''New methods in the statistical theory of nonequilibrium processes,'' in *Current Problems of Mathematics* [in Russian], Vol. 15, VINITI Akad. Nauk SSSR (1979).
[18] A. Erdélyi, *Higher Transcendental Functions* (Bateman Project) Vol. 2, McGraw-Hill, New York (1953).
[19] W. B. Jones and W. J. Thron, *Continued Fractions. Analytic Theory and Applications*, Mir Publishers, Moscow (1985) [English original: Addison-Wesley, Reading, Mass. (1980)].
[20] V. L. Berezinskiĭ and L. P. Gor'kov, Zh. Éksp. Teor. Fiz. **77**, 2498 (1979) [Sov. Phys. JETP **50**, 1209 (1979)].
[21] V. N. Prigodin and Yu. A. Firsov, J. Phys. C **17**, L979 (1984).
[22] Y. Imry, Y. Gefen, and D. J. Bergman, in *Anderson Localization*, Y. Nagaoka and H. Fukuyama (eds.), Springer, Berlin (1982), p. 138.
[23] H. F. Hess, K. DeConde, T. F. Rosenbaum, and G. A. Thomas, Phys. Rev. B **25**, 5578 (1982).
[24] É. Z. Kuchinskiĭ, M. V. Sadovskiĭ, V. G. Suvorov, and M. A. Érkabaev, Zh. Éksp. Teor. Fiz. **107**, 2027 (1995) [JETP **80**, 1122 (1995)].
[25] É. Z. Kuchinskiĭ and M. A. Érkabaev, Fiz. Tverd. Tela (St. Petersburg) **39**, 412 (1997) [Phys. Solid State **39**, 357 (1997)].
[26] H. Bateman and A. Erdélyi, *Tables of Integral Transforms*, Vol. 2, McGraw-Hill, New York (1954).
[27] A. Erdélyi, *Higher Transcendental Functions*, (Bateman Project) Vol. 1, McGraw-Hill, New York (1953).

Translated by Eugene Yankovsky

Nonlinear waves in zinc

V. G. Skobov and A. S. Chernov

Moscow State Institute of Engineering Physics, 115409 Moscow, Russia
(Submitted 27 February 1997)
Zh. Éksp. Teor. Fiz. **114**, 725–734 (August 1998)

A theoretical analysis is made of Doppler-shifted cyclotron resonance in zinc in linear and nonlinear regimes. It is shown that the absence of a threshold for cyclotron absorption by holes makes the doppleron strongly damped or eliminates it for low-amplitude wave-fields. At high amplitudes capture of holes by the magnetic field of the wave suppresses collisionless absorption and the doppleron can propagate. As a result, the impedance of the plate is an oscillating function of the magnetic field. It is shown that the effect should be observed at frequencies of the order of a few tens of kilohertz, in magnetic fields of the order of a few kilogauss, and for exciting field amplitudes of the order of a few tens of gauss. © *1998 American Institute of Physics.* [S1063-7761(98)02708-5]

1. INTRODUCTION

Numerous theoretical and experimental investigations have been made of dopplerons in cadmium caused by Doppler-shifted cyclotron resonance (DSCR) of lens electrons, and the main results are reviewed in Ref. 1. However, no dopplerons have been observed in zinc, whose Fermi surface contains an electron lens very similar to that of cadmium. This can be attributed to the difference between the Fermi surfaces of zinc and cadmium. The hole ''monster'' of zinc consists of six columns connected by ''arms'' distributed in the central plane of the Brillouin zone. In cadmium these arms are broken. This last feature means that the maximum displacement of the holes per cyclotron period in cadmium is several times smaller than the displacement of the electrons at the reference point of the lens, and in the wavelength range corresponding to electron DSCR no collisionless cyclotron absorption by holes is observed. Thus, an electron doppleron in cadmium undergoes no collisionless damping and its excitation gives rise to clearly defined oscillations of the plate impedance as a function of the magnetic field. Zinc however, has carrier orbits passing through the arms of the monster for which the displacements of the holes may have arbitrarily high values. In consequence, cyclotron absorption by holes occurs at any wavelengths, including those near DSCR of the lens electrons. This collisionless absorption makes a significant contribution to the damping of an electron doppleron, as a result of which this doppleron is not observed.

This is the scenario for low-amplitude rf exciting fields. At high amplitudes the situation is different. In this case, carriers may be trapped by the magnetic field of the wave, leading to a substantial drop in collisionless absorption. It was established in Ref. 2 that capture of electrons responsible for collisionless damping of a hole doppleron in cadmium substantially increases the amplitude of the corresponding doppleron oscillations of the plate impedance in the nonlinear regime. It was shown later in Refs. 3 and 4 that in noble metals, suppression of collisionless absorption by carriers with open orbits makes it possible for helicons to propagate in a geometry in which they are not observed in the linear regime ($\mathbf{k}||\mathbf{H}||[110]$, where $\mathbf{k}$ is the wave propagation vector and $\mathbf{H}$ is the static magnetic field). It was also shown in Refs. 5 and 4 that suppression of cyclotron absorption in noble metals should lead to the existence of a new nonlinear wave in comparatively weak magnetic fields (appreciably below the helicon threshold), which has no analog in the linear regime.

Here we show that cyclotron absorption by holes does in fact lead to strong damping of an electron doppleron in zinc in the linear regime and may be substantially suppressed at quite attainable amplitudes of the exciting field in the nonlinear regime. As a result, it becomes possible to observe a nonlinear electron doppleron in zinc.

2. MODEL OF THE FERMI SURFACE AND NONLOCAL CONDUCTIVITY

Here we consider DSCR in zinc in a $\mathbf{k}\|\mathbf{H}\|C_6$ geometry. A characteristic property of its hole Fermi surface is the presence of saddle points. These points lie in the plane perpendicular to the C_6 axis which separates the cross sections of the monster passing through the arms from the cross sections passing only through the columns. Two such planes exist, positioned symmetrically relative to the central plane, each having six saddle points lying on the upper (or lower) surfaces of arms exactly midway between neighboring columns. It is known that on orbits passing through the saddle points, the cyclotron mass and thus the displacement of the carriers per cyclotron period, become infinite. Since the orbits of holes positioned between two such planes differ negligibly from circular, this part of the monster may be approximated by an axisymmetric surface. We shall consider a model in which the cross-sectional plane of the hole Fermi surface S_h depends on the longitudinal component of the momentum p_z as follows:

$$S_h(p_z)=2\pi p p_2 \times\begin{cases} a+\sqrt{1-(p_z/p_2)^2}, & |p_z|\leqslant p_2, \\ a-\sqrt{1-(2p_2-|p_z|)^2/p_2^2}, & p_2\leqslant|p_z|\leqslant 2p_2, \end{cases} \tag{1}$$

where $a>1$ and p and p_2 are the dimensional parameters of the momentum. It follows from Eq. (1) that the displacement of the holes per cyclotron period, defined by the derivative $\partial S_h/\partial p_z$ has a root characteristic for $p_z=p_2$:

$$u_h=-\frac{c}{eH}\frac{\partial S_h}{\partial p_z}, \tag{2}$$

$$\frac{\partial S_h}{\partial p_z}=\begin{cases} -2\pi p\dfrac{p_z/p_2}{\sqrt{1-(p_z/p_2)^2}}, & |p_z|<p_2, \\ \dfrac{2\pi p(|p_z|-2p_2)\operatorname{sign} p_z}{p_2\sqrt{1-(2p_2-|p_z|)^2/p_2^2}}, & p_2<|p_z|<2p_2. \end{cases} \tag{3}$$

Thus, the parameter p determines the characteristic displacement of the holes. The value of p is of the order of the corresponding value for the electrons at the reference point of the lens. The hole concentration equal to the electron concentration is

$$n=\frac{2}{(2\pi\hbar)^3}\int_{-2p_2}^{2p_2} S_h(p_z)\,dp_z. \tag{4}$$

When the wave propagation vector and the static magnetic field are directed along the axis of the axisymmetric Fermi surface ($\mathbf{k}\|\mathbf{H}\|z$), the nonlocal conductivity is given by (see Ref. 1, for example):

$$\sigma_{\pm}(k,H)=\sigma_{xx}\pm\sigma_{yx} =\frac{ec}{H}\frac{2}{(2\pi\hbar)^3}\int\frac{S_h(p_z)\,dp_z}{\gamma+i[\pm 1-ku(p_z)/2\pi]}, \tag{5}$$

where e is the absolute value of the electron charge, c is the velocity of light, $\gamma=\nu/\omega_c$, ν is the frequency of collisions between carriers and scatterers, $\omega_c=eH/mc$ is the cyclotron frequency of the carriers, and m is their cyclotron mass. In Eq. (5) we neglected the dependence of $\sigma_{\pm}$ on the wave frequency ω, assuming that $\omega\ll\nu$.

We shall derive an expression for the nonlocal conductivity which determines the properties of a doppleron produced by the DSCR of the lens electrons, which is circularly polarized with a minus sign. Substituting Eqs. (1)–(3) into Eq. (5) and integrating over p_z with allowance for Eq. (4) yields the following expression for the hole component of the conductivity:

$$\sigma_-^h(k,H)=i\frac{nec}{H}\frac{1}{I_h^2+q_h^2} \times\left[I_h+\frac{q_h^2}{2\sqrt{I_h^2+q_h^2}}\left(\ln\frac{\sqrt{I_h^2+q_h^2}+I_h}{\sqrt{I_h^2+q_h^2}-I_h}-i\pi\right)\right], \tag{6}$$

where

$$q_h=kcp/eH,\quad I_h=1+i\gamma_h, \tag{7}$$

$\gamma_h=\nu_h/\omega_{ch}$, ω_{ch} and ν_h are the cyclotron frequency and collision frequency of the holes. Here we are interested in the situation where the scattering of carriers is weak and the magnetic field is strong, such that $\gamma_h\ll 1$.

The component σ_-^h, proportional to $-i\pi$ in parentheses in Eq. (6) describes the collisionless absorption of the wave by those holes whose displacement over the cyclotron period u_h is equal to the wavelength $2\pi/k$. This cyclotron absorption has no threshold in terms of q_h since in the range of p_z between $-p_2$ and p_2, the derivative $\partial S/\partial p_z$ varies between $-\infty$ and ∞, so that for any wavelength $2\pi/k$ there exist holes for which the condition $u_h(p_z)=2\pi/k$ is satisfied. Thus, the presence of saddle points on the Fermi surface means that cyclotron absorption exists in the metal at any wavelengths.

We are subsequently interested in the vicinity of the DSCR of the lens electrons where the value of q_h is slightly less than unity. In this case, the dissipative component σ_-^h caused by scattering of holes is much smaller than the term associated with cyclotron absorption. Thus, we can neglect terms of order γ_h and set $I_h=1$ so that the expression for σ_-^h has the form

$$\sigma_-^h=i\frac{nec}{H}G(q_h),$$

$$G(q_h)=\frac{1}{1+q_h^2}+\frac{q_h^2}{2(1+q_h^2)^{3/2}}\left(\ln\frac{\sqrt{1+q_h^2}+1}{\sqrt{1+q_h^2}-1}-i\pi\right). \tag{8}$$

In order to describe the electron lens, we shall use a model which we proposed in Ref. 6 in an analysis of DSCR in cadmium. In this model, the dependence of the cross-sectional area of the lens on the longitudinal electron momentum has the form

$$S_e(p_z)=4p_0p_3\left(\cos\frac{\pi p_z}{2p_3}-\rho\right), \tag{9}$$

where ρ is a dimensionless parameter and p_0 and p_3 are the dimensional parameters of the momentum. For $\rho\geqslant -1$ Eq. (9) describes a closed convex body extending along p_z from $-p_F$ to p_F, where

$$p_F=\frac{2p_3}{\pi}\arccos\rho. \tag{10}$$

The singularity of the nonlocal electron conductivity determined by the form of the function $\partial S_e/\partial p_z$ depends on the value of ρ. In the range $-1\leqslant\rho<0$ the conductivity has a root singularity, for $\rho=0$ it is logarithmic, and for $\rho>0$ the singularity becomes weakly logarithmic. We shall subsequently analyze electron DSCR in two cases: $\rho=-1$ and $\rho>0$. The case $\rho=-1$ is the most clear-cut: the nonlocal electron conductivity has a particularly simple form and the problem of calculating the impedance of the plate can be completely resolved analytically. Substituting Eq. (9) into

Eq. (5) for $\rho=-1$ and integrating over p_z allowing for Eq. (10) yields the following expression for the electron conductivity:

$$\sigma_-^e(k,H)=-i\frac{nec}{H}\frac{1}{\sqrt{I^2-q^2}}, \tag{11}$$

where

$$q=kp_0c/eH, \quad I=-1+i\gamma, \tag{12}$$

$$n=4p_0p_3^2/(\pi\hbar)^3, \tag{13}$$

n is the concentration of lens electrons.

3. DISPERSION EQUATION

The dispersion equation for a negatively polarized rf wave has the form

$$k^2c^2=4\pi i\omega[\sigma_-^e(k,H)+\sigma_-^h(k,H)]. \tag{14}$$

The root of the dispersion equation corresponding to an electron doppleron lies in the range $q^2<1$. An analysis shows that in this range the dependence of the nondissipative component σ_-^h on k is considerably weaker than σ_-^e. Thus, when solving the dispersion equation, we can confine ourselves to allowance for $\mathrm{Im}\,\sigma_-^h$ in the local approximation and for $\mathrm{Re}\,\sigma_-^h$ we need only retain the quadratic term in k^2. The dispersion equation can then be written in the form

$$D(q)=0, \tag{15}$$

where

$$D(q)=q^2-\xi_0\left(\frac{1}{\sqrt{I^2-q^2}}-1+i\alpha q^2\right), \tag{16}$$

$$\xi_0=4\pi\omega np_0^2c/eH^3, \tag{17}$$

$$\alpha=(\pi/2)(p^2/p_0^2), \tag{18}$$

the term -1 in parentheses in Eq. (16) corresponds to the local Hall conductivity of the holes and the last term is associated with cyclotron absorption.

Equation (15) can be transformed to give a cubic equation for q^2 whose roots have the form

$$\frac{q_l^2}{I^2}=\frac{1}{3}(1-2I\xi)+\frac{2}{3}(1+I\xi)\times\sin\left\{\frac{2\pi l}{3}-\frac{1}{3}\arcsin\left[1-\frac{27}{2}\frac{\xi^2}{(1+I\xi)^3}\right]\right\}, \quad l=0,1,2, \tag{19}$$

where

$$\xi=\xi_0/(1-i\alpha\xi_0). \tag{20}$$

We shall subsequently take q_l to mean those roots positioned in the upper half-plane of the complex plane q. In the range of fields H above the doppleron threshold ($\xi_0<2$) we find $q_0\sim\sqrt{i\gamma}$. This small complex root applies to the damped skin component of the wave field. The root q_1 refers to a doppleron and its value tends asymptotically to -1 in strong fields ($\xi_0\ll1$). The root q_2 is located on the second sheet of the Riemann surface of the function $D(q)$ and does not correspond to any component of the wave field.

4. IMPEDANCE OF THE PLATE

The impedance of a compensated metal plate with a root singularity of the nonlocal conductivity was studied theoretically in Ref. 7. An analysis was made of oscillations of the impedance caused by the excitation of a propagating wave mode (a doppleron), and of Gantmakher–Kaner oscillations caused by the excitation of a nonexponential component of the wave field as a result of the presence of a branching point ($q^2=I^2$) in the nonlocal conductivity (11). It was shown that if the thickness of the plate is not too small, the amplitude of the doppleron oscillations substantially exceeds that of the Gantmakher–Kaner oscillations. Since we are interested in the possible propagation of a doppleron in zinc, we shall subsequently analyze this situation and in the expression for the plate impedance we shall neglect the term describing the Gantmakher–Kaner oscillations. The expression for the impedance of the plate can then be expressed in the form

$$Z=Z_s[1-\kappa Z_sb^2D'(q_1)A\exp(iq_1L)], \tag{21}$$

$$(\kappa Z_s)^{-1}=q_0\frac{1+\exp(iq_0L)}{1-\exp(iq_0L)}+q_1-\frac{I}{2}-\frac{1}{\pi}\times\left(q_0\arcsin\frac{q_0}{I}+q_1\arcsin\frac{q_1}{I}-q_2\arcsin\frac{q_2}{I}\right), \tag{22}$$

where

$$b=-2q_1(q_0+q_1)\exp[J(q_1)]/D'(q_1), \quad D'(q)=dD(q)/dq, \tag{23}$$

$$J(q)=-\frac{1}{2}\ln\left(1+\frac{q}{I}\right)-\frac{\xi_0}{\pi}\times\int_0^\infty\frac{(3z^2+1+\xi_0)\ln[(1+z^2)^{1/2}+q/I]dz}{(1+z^2)[(z^2+\xi_0)^2+z^2]}, \tag{24}$$

$$A=\exp\left[-\left(\frac{seH}{p_0c}\right)^2\right], \tag{25}$$

$$\kappa=\frac{ceH}{8\pi\omega p_0}, \quad L=\frac{eH}{cp_0}d, \tag{26}$$

d is the plate thickness and s is the mean-square dimension of the surface roughness of the plate. The quantity Z_s represents the smooth part of the plate impedance and the term containing the factor $\exp(iq_1L)$ describes the oscillations of the impedance as a function of H caused by the excitation of a doppleron wave in the plate. The factor A in Eq. (21) describes the decrease in the amplitude of the doppleron oscillations as a result of the surface roughness.[6]

In the range of strong fields $\xi\ll1$ the integral term in Eq. (24) is equal to the first term, and as H decreases to the doppleron threshold H_L it varies by less than 10%. Thus, $J(q_1)$ may be approximated by

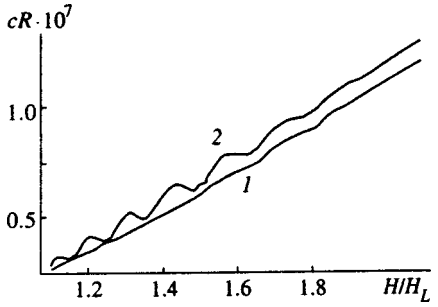


FIG. 1. Surface resistance of the plate versus magnetic field in linear (curve *1*) and nonlinear (curve *2*) regimes with $\omega/2\pi=100$ kHz and $d=2$ mm ($H_L=3$ kOe).

$$J(q_1)\simeq -\ln(1+q_1/I)\,, \tag{27}$$

and as a result, expression (23) for b is simplified and has the form

$$b=-\frac{2q_1(q_0+q_1)}{(1+q_1/I)D'(q_1)}. \tag{28}$$

Figure 1 gives the results of calculations of the surface resistance $R=\mathrm{Re}\,Z$ of the plate as a function of H for a low-amplitude rf exciting field when linear theory applies (curve *1*). The calculations are made for

$$p_0=1.5\ \hbar\ \text{Å}^{-1},\quad n=5\times10^{21}\,\text{cm}^{-3}\,,\quad \nu=4\times10^{8}\,\text{s}^{-1}\,, \tag{29}$$

exciting field frequency $\omega/2\pi=100$ kHz, plate thickness $d=2$ mm, and $s=3\times10^{-3}$ cm. It can be seen that the graph $R(H)$ is a monotonically increasing function with almost no oscillations. This is attributable to two factors. First, doppleron oscillations are suppressed as a result of the roughness of the plate: in strong magnetic fields H where the displacement of the resonant electrons $2\pi cp_0/eH$ becomes comparable with the size of the roughness, the value of A is much less than unity. Second, cyclotron absorption by holes substantially reduces the doppleron damping length w. In Fig. 2 the solid curve gives $w(H)$. For comparison the dashed curve gives $w_0(H)$ in the absence of cyclotron absorption by holes ($\alpha=0$). It can be seen that in moderate magnetic fields $H>H_L$ the value of w is considerably less than the plate thickness d. As a result, a doppleron excited by an external field at one surface of the plate does not reach the other surface. Thus, in strong magnetic fields the oscillations are suppressed as a result of the roughness of the plate while in weak fields they are suppressed as a result of collisionless damping.

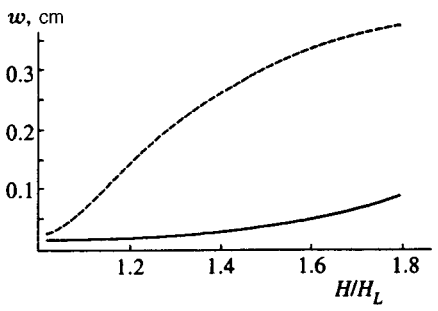


FIG. 2. Doppleron damping length versus static magnetic field in linear (solid curve) and nonlinear (dashed curve) regimes.

In the nonlinear regime the situation may change substantially. The magnetic field of a large-amplitude wave ''captures'' holes with $p_z=p_z^0$ which satisfy the condition

$$u_h(p_z^0)=2\pi/k. \tag{30}$$

The longitudinal velocity of these holes is modulated at the frequency[2]

$$\omega_0=\frac{e}{mc}\left|HH_a\left(\frac{|S_h|}{\pi}\right)^{1/2}\frac{\partial^2 S_h}{\partial p_z^2}\Big/\frac{\partial S_h}{\partial p_z}\right|^{1/2}_{p_z=p_z^0}, \tag{31}$$

where H_a is the amplitude of the magnetic field of the wave in the metal. For our model of a hole Fermi surface we have

$$\omega_0\sim\omega_{ch}\sqrt{H_a/H}. \tag{32}$$

In the nonlinear regime when the frequency ω_0 is much greater than the hole collision frequency ν_h, cyclotron absorption decreases as the ratio ν_h/ω_0 decreases. This implies that the substitution $\alpha\to\alpha\nu_h/\omega_0$ must be made in expression (16) for $D(q)$. As a result, the doppleron damping is reduced. If the ratio ω_0/ν_h is fairly large, the doppleron damping length increases from w to w_0, as is shown by the dashed curve in Fig. 2. In this case, the dependence $R(H)$ is given by curve *2* in Fig. 1. Unlike curve *1*, curve *2* clearly reveals doppleron oscillations.

5. DOPPLER-SHIFTED CYCLOTRON RESONANCE OF REFERENCE-POINT ELECTRONS

We have so far considered the case when in which the DSCR is caused by electrons with an orbit of finite dimensions and the conductivity has a root singularity. However, it is more likely that DSCR in zinc is caused by electrons at the reference point of the lens and that the conductivity σ^e_- has a weak logarithmic singularity. We shall now analyze this case. We shall assume that the parameter ρ which characterizes the shape of the lens and the type of resonance, is a small positive quantity. In order to ensure that the highest value of $|\partial S_e/\partial p_z|$ is equal to $2\pi p_0$ as before, we write the expression for $S_e(p_z)$ not in the form (9) but in a slightly different form:

$$S_e(p_z)=2\pi p_0\frac{p_1}{\sigma^2}\left(\cos\frac{\sigma p_z}{p_1}-\rho\right),\quad |p_z|\leqslant p_F\,, \tag{33}$$

$$\sigma=\sqrt{1-\rho^2},\quad p_F=(p_1/\sigma)\arcsin\sigma, \tag{34}$$

where p_1 is a parameter with dimensions of momentum.

The longitudinal electron velocity is

$$v_{ze}=-\frac{1}{2\pi m_e}\frac{\partial S_e}{\partial p_z}=\frac{p_0}{m_e\sigma}\sin\frac{\sigma p_z}{p_1}, \tag{35}$$

its maximum is achieved at the reference point of the lens, $p_z=p_F$, and equals $v_0=p_0/m_e$. The electron concentration n is now determined by

$$n=\frac{p_0 p_1^2}{\pi^2\sigma^2\hbar^3}\lambda, \quad \lambda=1-\rho\frac{p_F}{p_1}. \tag{36}$$

Substituting Eq. (33) into Eq. (5) and integrating over p_z gives

$$\sigma_-^e(k,H)=-i\frac{nec}{H}f(q), \tag{37}$$

$$f(q)=\frac{1}{\lambda}\left[\frac{1}{2q}\ln\frac{I-q}{I+q}+i\frac{\rho}{2\sqrt{\sigma^2I^2-q^2}}\ln\frac{\rho+i\sqrt{\sigma^2-q^2/I^2}}{\rho-\sqrt{\sigma^2-q^2/I^2}}\right], \tag{38}$$

where q and I are given by the formulas (12).

The expression for the function $D(q)$ on the left-hand side of the dispersion equation (15) now has the form

$$D(q)=q^2-\xi_0[f(q)-G(\eta q)], \quad \eta=p/p_0, \tag{39}$$

where G is given by the formula (8).

In this model the region of existence of the doppleron not only has a lower threshold in terms of H but also an upper threshold.[6] The lower threshold H_L is given by

$$H_L=\left(\frac{4\pi\omega n p_0^2 c}{e\xi_L}\right)^{1/3}, \quad \xi_L=6\left(\frac{3}{\sigma^2}-\frac{1}{\lambda}\right)^{-1}. \tag{40}$$

The upper threshold H_U corresponding to the condition $q^2=1$ is obtained from the equation $D(1)=0$ and is given by

$$H_U=H_L\left(\frac{\xi_L}{\xi_U}\right)^{1/3}, \quad \xi_U=\left(\frac{1}{\lambda}\ln\frac{1}{\rho}-1\right)^{-1}. \tag{41}$$

It follows from Eq. (41) that H_U increases as ρ decreases. For $\rho\to 0$ the value of H_U increases without bound, which corresponds to the conversion of an elliptic reference point into a parabolic one.

In this model, the dispersion equation $D(q)=0$, as before, has the skin root q_0 and the doppleron root q_1 although these can now only be found numerically. Thus, the smooth component of the plate impedance is given by the integral

$$(\kappa Z_s)^{-1}=q_0\frac{1+\exp(iq_0L)}{1-\exp(iq_0L)}+q_1+K, \tag{42}$$

$$K=\frac{\xi_0}{\pi}\int_1^\infty q\,dq\,\frac{2q\Delta-q^2\Delta'-\xi_0(F'\Delta-F\Delta')}{(q^2-\xi_0F)^2+\xi_0^2\Delta^2}, \tag{43}$$

where

$$F(q)=\frac{1}{\lambda}\left[\frac{1}{2q}\ln\frac{q-I}{q+I}-\frac{\rho}{2\sqrt{q^2-\sigma^2I^2}}\times\ln\frac{\sqrt{q^2/I^2-\sigma^2}+\rho}{\sqrt{q^2/I^2-\sigma^2}-\rho}\right]-G(\eta q), \tag{44}$$

$$\Delta(q)=\frac{\pi}{2\lambda}\left(\frac{1}{q}-\frac{\rho}{\sqrt{q^2-\sigma^2I^2}}\right), \tag{45}$$

and the prime indicates differentiation with respect to q. We also note that as in the model with a root singularity, the value of b characterizing the amplitude of the doppleron oscillations may be approximated by the expression (28).

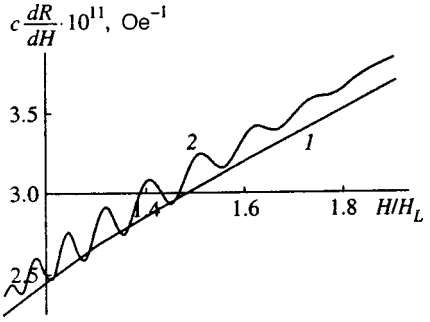


FIG. 3. Graphs of the derivative of the surface resistance of the plate with respect to the magnetic field in the linear (curve *1*) and nonlinear (curve *2*) regimes with $\omega/2\pi=100$ kHz and $d=1.5$ mm ($H_L=3$ kOe).

The results of calculations of the derivative of the surface resistance of the plate dR/dH for the values of the parameters (29), frequency $\omega/2\pi=100$ kHz, thickness $d=1.5$ mm, and $s=4\times10^{-3}$ cm are plotted in Fig. 3. Curve *1* corresponds to the linear regime and curve *2* to the nonlinear. Curve *2* was obtained for a wave field amplitude $H_a=20$ G, for which the ratio ν/ω_0 is less than 0.1 in the region where dopplerons exist. Thus for moderate values of H_a the suppression of cyclotron absorption is so great that it is possible for a doppleron to propagate.

In these calculations, the parameter α given by Eq. (18), which determines the cyclotron damping, is close to unity. In practice, existing information on the hole monster in zinc is inadequate for an accurate determination of the parameter p contained in formula (18) for α. The real value of α may be larger than we assumed. It is therefore advisable to analyze how the value of α influences the properties of the doppleron. The analysis shows that as α increases, the doppleron root q_1 of the dispersion equation passes counterclockwise around the conductivity branch point $q=I$ and at a certain value $\alpha=\alpha_0$ is transferred below the cut to a nonphysical sheet of the Riemann surface of the function $D(q)$. This critical value of α_0 depends on the magnetic field H. It has a maximum near the range of existence of the doppleron and decreases as H approaches the upper or lower threshold. As a result, it is found that as α increases, the upper and lower threshold fields H_U and H_L converge and for $\alpha\simeq 2$ the doppleron disappears. Thus, if the monster in zinc is such that $\alpha>2$, the nonlinear effect will be more striking: in the nonlinear regime a doppleron can propagate whereas in the linear regime it does not generally exist.

[1] A. S. Chernov and V. G. Skobov, Phys. Rep. **244**, 1 (1994).
[2] I. F. Voloshin, G. A. Vugal'ter, V. Ya. Demikhovskiĭ *et al.*, Zh. Éksp. Teor. Fiz. **72**, 1503 (1977) [Sov. Phys. JETP **46**, 790 (1977)].
[3] V. G. Skobov and A. S. Chernov, JETP Lett. **61**, 1012 (1995).
[4] V. G. Skobov and A. S. Chernov, Zh. Éksp. Teor. Fiz. **109**, 992 (1996) [JETP **82**, 535 (1996)].
[5] A. S. Chernov and V. G. Skobov, Phys. Lett. A **205**, 81 (1995).
[6] N. A. Podlevskikh, V. G. Skobov, L. M. Fisher, and A. S. Chernov, Fiz. Tverd. Tela **27**, 330 (1985) [Sov. Phys. Solid State **27**, 202 (1985)].
[7] I. F. Voloshin, V. G. Skobov, L. M. Fisher, and A. S. Chernov, Zh. Éksp. Teor. Fiz. **82**, 293 (1982) [Sov. Phys. JETP **55**, 175 (1982)].

Translated by R. M. Durham

Magnetic soliton motion in a nonuniform magnetic field

A. M. Kosevich*)

Low-Temperature Physicotechnical Institute, National Academy of Sciences, 310086 Kharkov, Ukraine

V. V. Gann and A. I. Zhukov

National Research Center ''Kharkov Physicotechnical Institute,'' 310108 Kharkov, Ukraine

V. P. Voronov

Belgorod State University, 308000 Belgorod, Russia

(Submitted 12 December 1997)
Zh. Éksp. Teor. Fiz. **114**, 735–746 (August 1998)

We discuss the dynamics of a magnetic soliton in a one-dimensional ferromagnet placed in a weakly nonuniform magnetic field. In the presence of a constant weak magnetic-field gradient the soliton quasimomentum is a linear function of time, which induces oscillatory motion of the soliton with a frequency determined by the magnetic-field gradient; the phenomenon is similar to Bloch oscillations of an electron in a weak electric field. An explicit description of soliton oscillations in the presence of a weak magnetic-field gradient is given in the adiabatic approximation. Two turning points are found in the motion of the soliton and the varieties of bounded and unbounded soliton motion are discussed. The Landau–Lifshitz equations are solved numerically for the case of a soliton moving in a weakly nonuniform magnetic field. The soliton is shown to emit a low-intensity spin wave near one of the turning points due to violation of the adiabatic approximation, and the necessary conditions for such an approximation to hold are established. [S1063-7761(98)02808-X]

1. INTRODUCTION

A remarkable feature of the nonlinear dynamics of the magnetization of ferromagnets and antiferromagnets is the presence of dynamic magnetic solitons. By a dynamic soliton we mean a spatially localized perturbation in the magnetization field whose stability is ensured by the presence of certain integrals of motion for the dynamical equations of this field. In one-dimensional ferromagnets, where the magnetization dynamics is described by Landau–Lifshitz equations, a complete description of all types of nonlinear excitations is possible; in particular, there exists an exact analytical description of dynamic solitons in uniaxial and biaxial magnetic materials in a uniform magnetic field in the absence of dissipation.[1,2]

The phenomenological Landau–Lifshitz equations for ferromagnets have a quantum mechanical basis and provide an accurate explanation of the dynamical properties of magnetically ordered media. This suggests the possibility of broadening the range of physical phenomena to which the Landau–Lifshitz equations can be applied. In particular, we will describe the important effect of nonuniformity of the magnetic field on the motion of a magnetic soliton. To formulate the problem within a general setting, we start by recalling the basic principles of the nonlinear dynamics of the magnetization of ferromagnets at low temperatures.

The instantaneous state of a ferromagnet is determined by the magnetization vector as a function of position and time, $\mathbf{M}(\mathbf{r},t)$. According to current ideas about the exchange spin nature of ferromagnetism, the magnitude M_0 of the magnetization vector remains unchanged, so that the magnetization dynamics reduces to the precessional motion of this vector.[3] In other words, if we introduce the polar angles θ and φ, the magnetization vector $\mathbf{M}$ of the ferromagnet can be written as

$$M_x + iM_y = M_0 \sin\theta e^{i\varphi}, \quad M_z = M_0 \cos\theta. \tag{1}$$

In terms of the angular variables θ and φ, the Landau–Lifshitz equations

$$\frac{\partial \mathbf{M}}{\partial t} = -\frac{2\mu_0}{\hbar}\,\mathbf{M}\times\frac{\delta E}{\delta \mathbf{M}}$$

take the form

$$\sin\theta\,\frac{\partial\theta}{\partial t} = -\frac{2\mu_0}{\hbar M_0}\,\frac{\delta E}{\delta\varphi}, \quad \sin\theta\,\frac{\partial\varphi}{\partial t} = \frac{2\mu_0}{\hbar M_0}\,\frac{\delta E}{\delta\theta}, \tag{2}$$

where the right-hand sides of the equations contain variational derivatives of the total magnetic energy of the magnetic material, E, with respect to the magnetization and the angular variables, and μ_0 is the Bohr magneton. The total energy E can be written as

$$E = \int w\{\theta,\varphi\}\,d^3x, \tag{3}$$

where the magnetic energy density w depends on the angular variables θ and φ and their gradients.

We limit our discussion to the case of a ferromagnet with uniaxial magnetic anisotropy placed in an external mag-

netic field **H** that is directed along the anisotropy axis **n**. We identify the z axis with this axis. Then the magnetic energy density can be written as[3]

$$w\{\theta,\varphi\}=w_0(\theta,\nabla\theta,\nabla\varphi)+M_0(1-\cos\theta)H, \tag{4}$$

where

$$w_0=\frac{1}{2}\alpha\left(\frac{\partial\mathbf{M}}{\partial x_i}\right)^2+\frac{1}{2}\beta M_0^2\sin^2\theta,$$

with α being the exchange constant and β the anisotropy constant. The function $w_0(\theta,\nabla\theta,\nabla\varphi)$ depends on the gradients of the angular variables but does not depend explicitly on the angular variable φ (the phase). Some statements referring to a magnetic soliton are unrelated to the specific form of the function w_0 if the latter depends on the specified arguments.

The Landau–Lifshitz equations for a uniaxial ferromagnet always has two constants of motion: the total magnetic-excitation energy E, and the projection of the total magnetic moment on the anisotropy axis. The second constant of motion, related to the presence of the cyclic coordinate φ, can be conveniently written as

$$N=\frac{1}{2\mu_0}\int[M_0-M_z(\theta)]d^3x$$
$$=\frac{M_0}{2\mu_0}\int(1-\cos\theta)d^3x. \tag{5}$$

The normalization (5) makes it possible to assume that N is the number of magnons whose bound state forms the soliton.[1,2]

If the external magnetic field is uniform, the total excitation field momentum (total quasimomentum)

$$\mathbf{P}=-\frac{\hbar M_0}{2\mu_0}\int(1-\cos\theta)\nabla\varphi d^3x \tag{6}$$

is also conserved (in addition to E and N).

A dynamic magnetic soliton is a solution of Eqs. (2) localized in space, moving with a constant velocity, and corresponding to finite values of the constants of motions E, N, and **P**. Such a solution has the form

$$\theta=\theta(\mathbf{r}-\mathbf{V}t),\quad \varphi=\Omega t+\psi(\mathbf{r}-\mathbf{V}t), \tag{7}$$

where **V** is the soliton velocity, Ω is the frequency of the soliton's internal precession, and the functions $\theta(\xi)$ and $\varphi(\xi)$ possess the following properties:

$$\theta(\xi)=0,\quad |\nabla\varphi|<\infty\quad \text{as}\quad \xi\to\pm\infty. \tag{8}$$

Hence a magnetic soliton is a two-parameter excitation, with **V** and Ω being the parameters.

The constants of motion E, N, and **P** are connected by a remarkable relationship, which is independent of the type of the functions (7); namely, under small variations of the functions θ and φ the variation of the total energy is[1,2]

$$\delta E=\mathbf{V}\cdot\delta\mathbf{P}+\hbar\Omega\,\delta N. \tag{9}$$

This yields two equations of motion for the soliton:

$$\mathbf{V}=\left(\frac{\partial E}{\partial\mathbf{P}}\right)_N,\quad \hbar\Omega=\left(\frac{\partial E}{\partial N}\right)_{\mathbf{P}}, \tag{10}$$

where the first determines the rate of variation of the position of the soliton's center of gravity, and the second the rate of variation of its phase. In a uniaxial ferromagnet in the presence of a uniform magnetic field, the position of the soliton's center of gravity and its phase are cyclic variables, which ensure the validity of the following conservation laws:

$$\mathbf{P}=\text{const},\quad N=\text{const}. \tag{11}$$

In this paper we study the dynamics of a magnetic soliton in a uniaxial ferromagnet in a weakly nonuniform magnetic field; in particular, we investigate the case of a constant magnetic-field gradient:

$$H=H_0+\eta x,\quad \eta=\frac{dH}{dx}. \tag{12}$$

In Sec. 2 we show that in this case the soliton position ceases to be a cyclic variable, with the result that the quasimomentum **P** ceases to be a constant of motion. When η is small, the quasimomentum **P** is a linear function of time, which changes the soliton dynamics dramatically.

For an object for which the calculations can be carried out analytically we take a one-dimensional easy-axis ferromagnet. The magnetic energy density w_0 of such a ferromagnet in terms of the angular variables θ and φ is

$$w_0=\frac{1}{2}\alpha M_0^2\left[\left(\frac{\partial\theta}{\partial x}\right)^2+\sin^2\theta\left(\frac{\partial\varphi}{\partial x}\right)^2\right]+\frac{1}{2}\beta M_0^2\sin^2\theta. \tag{13}$$

We discuss the dynamics of a magnetic soliton with the magnetic energy (13) in a one-dimensional uniaxial ferromagnet. In a uniform magnetic field, the energy of such a soliton is a periodic function of P (Refs. 1 and 2):

$$E=E_0(P,N)+2\mu_0NH_0,\quad E_0(P,N)=2W_0l_0\kappa(P,N),$$
$$l_0\kappa(P,N)=\tanh\frac{N}{N_1}\left[1+\frac{\sin^2(\pi P/2P_0)}{\sinh^2(N/N_1)}\right], \tag{14}$$

where $W_0=2M_0^2\sqrt{\alpha\beta}a^2$ is the surface energy of the domain boundary; $l_0=\sqrt{\alpha/\beta}\gg a$, with l_0 the characteristic magnetic length and a the interatomic separation; $P_0=\pi\hbar a^3M_0/\mu_0$; and $N_1=2a^2l_0M_0/\mu_0$ (here N_1 coincides in order of magnitude with the maximum number of spin deviations that can occur over the length l_0, and $P_0=2\pi s\hbar/a$, where s is the atomic spin, which determines the magnetism of the material).

If the magnetic-field gradient is so weak that $\eta l_0\ll H$, then in the presence of a small η the dependence of the energy E on P can still be described by Eq. (14) but the field momentum P must be assumed to be a linear function of time (this corresponds to what is known as the adiabatic approximation).

If the momentum P is a linear function of time, Eqs. (14) and (10) imply that the soliton oscillates with a frequency determined by the magnetic-field gradient. Since such motion is similar to the oscillations of a Bloch electron in a uniform electric field, we call it Bloch oscillations. The

phenomenon of Bloch oscillations of a magnetic soliton in a magnetic field with a weak gradient was noted recently by one of the present authors in Ref. 4.

In Sec. 3 we study the dynamics of a one-dimensional soliton in the more general case of a weakly nonuniform magnetic field. If the nonuniformity of the magnetic field is located in a bounded interval of the one-dimensional magnetic material (the x axis), it creates an effective potential barrier for the soliton's motion. We discuss the various variants of bounded and unbounded soliton motion in the presence of different potential barriers.

Finally, in Sec. 4 we present the results of a numerical solution of the one-dimensional Landau–Lifshitz equations for the case of a soliton moving in a weakly nonuniform magnetic field. We find that when a soliton is in oscillatory motion, near one of the turning point a low-intensity spin wave with a frequency Ω is emitted. Such emission of a spin wave is due to the violation of the condition of validity of the adiabatic approximation. The criteria for the applicability of such an approximation are discussed.

2. BLOCH OSCILLATIONS OF A SOLITON IN A MAGNETIC FIELD WITH A CONSTANT GRADIENT

Let us study the soliton dynamics of a one-dimensional uniaxial ferromagnet with uniaxial anisotropy placed in a nonuniform magnetic with a weak gradient of type (12). As noted earlier, in a nonuniform magnetic field, P ceases to be a constant of motion. Let us examine the emerging time dependence of the quasimomentum defined according to the definition (6) and the property (8):

$$\frac{dP}{dt}=\frac{\hbar M_0}{2\mu_0}\int\left[\sin\,\theta\,\frac{\partial\varphi}{\partial t}\,\frac{\partial\theta}{\partial x}-\sin\,\theta\,\frac{\partial\varphi}{\partial x}\,\frac{\partial\theta}{\partial t}\right]dx. \tag{15}$$

We now use the equations of motion (2):

$$\frac{dP}{dt}=\int\left[\frac{\delta E}{\delta\theta}\,\frac{\partial\theta}{\partial x}+\frac{\delta E}{\delta\varphi}\,\frac{\partial\varphi}{\partial x}\right]dx. \tag{16}$$

Since the magnetic field depends on the x coordinate, the magnetic energy density also depends on x explicitly: $w\{\theta,\varphi;x\}$. This means that

$$\frac{dP}{dt}=\int\left[dw\{\theta,\varphi;x\}-\frac{\partial w}{\partial x}\,dx\right]$$

$$=-\int\frac{\partial w}{\partial x}\,dx=-\eta M_0\int(1-\cos\,\theta)dx=-2\eta\mu_0 N. \tag{17}$$

Thus, the quasimomentum is a linear function of time:

$$P(t)=P(0)-2\eta\mu_0 Nt,\quad P(0)=\text{const}. \tag{18}$$

Let us now calculate the total energy E of a soliton moving in the field of the weak gradient η. A weak gradient of the magnetic field can be interpreted as a weak perturbation of the soliton's motion in a uniform field. Then, in the adiabatic approximation,[5–7] the soliton retains its shape and the distribution of magnetization in it remains the same function of x given by Eq. (7):

$$\theta=\theta(x-X(t)),\quad \varphi=\varphi_0(t)+\psi(x-X(t)), \tag{19}$$

where the coordinate $X(t)$ of the soliton's center of gravity and its phase $\varphi_0(t)$ are functions of time to be determined. The main dynamical parameters of the soliton, the velocity V of the center of gravity and the frequency Ω, are given by obvious relationships:

$$V=\frac{dX}{dt},\quad \Omega=\frac{d\varphi_0}{dt}. \tag{20}$$

The total soliton energy can be written as

$$E=E_0(P,N)+2\mu_0 NH_0+\eta M_0\int(1-\cos\,\theta)x\,dx, \tag{21}$$

where $E_0(P,N)$ is defined in (14), and in the third term on the right-hand side we must bear in mind that in the adiabatic approximation $\theta(\xi)=\theta(-\xi)$:

$$\int(1-\cos\,\theta)x\,dx=\int[1-\cos\,\theta(x-X(t))]x\,dx=X(t)\times\int(1-\cos\,\theta(\xi))d\xi=2\eta\mu_0 NX. \tag{22}$$

Combining (21) and (22), we get

$$E=E_0(P,N)+2\mu_0 NH_0+2\eta\mu_0 NX. \tag{23}$$

We see that the energy E is a function of three dynamical variables, P, X, and N, and that the expression (17) for the time derivative of the momentum serves as one of the canonical Hamiltonian equations:

$$\frac{dP}{dt}=-\frac{\partial E}{\partial X},\quad \frac{dX}{dt}=\frac{\partial E}{\partial P}. \tag{24}$$

On the other hand, selecting the initial coordinate of the soliton appropriately, we can use (23) and (14) to find the explicit time dependence of the coordinate of the soliton's center of gravity:

$$X(t)=X(0)+\frac{W_0[\cos(\pi P(t)/P_0)-\cos(\pi P(0)/P_0)]}{\eta\mu_0 N\,\sinh(2N/N_1)}, \tag{25}$$

where $P(t)$ is given in (18). If $P(0)\neq 0$ holds for short times, as long as $\eta\mu_0 Nt\ll P_0$ is satisfied, the soliton is in uniform motion:

$$X(t)=X(0)+\frac{2\pi W_0\,\sin(\pi P(0)/P_0)}{P_0\,\sinh(2N/N_1)}\,t. \tag{26}$$

For long times ($\eta\mu_0 Nt>P_0$), the soliton is in oscillatory motion. As Eq. (25) implies, the amplitude of the spatial oscillations is

$$\Delta X=\frac{W_0}{\eta\mu_0 N\,\sinh(2N/N_1)}. \tag{27}$$

Naturally, it is inversely proportional to the magnetic- field gradient and drops off rapidly as N increases, i.e., as the soliton grows in size:

$$V(t)\equiv\frac{dX}{dt}=V_m\,\frac{\sin(\pi P(t)/P_0)}{\sinh(2N/N_1)},\tag{28}$$

where $V_m=2gM_0\sqrt{\alpha\beta}$ is the minimum phase velocity of the spin waves ($g=2\mu_0/\hbar$).

The oscillations of the soliton precession frequency can be found from the second formula in (10) and the definition (23):

$$\Omega(t)=\frac{1}{\hbar}\,\frac{\partial E_0(P,N)}{\partial N}+gH_0+g\,\eta X(t),\tag{29}$$

with the first term on the right-hand side specified as

$$\frac{\partial E_0(P,N)}{\partial N}=\hbar\,\omega_0\left\{\frac{\cos^2(\pi P(t)/2P_0)}{\cosh^2(N/N_1)}-\frac{\sin^2(\pi P(t)/2P_0)}{\sinh^2(N/N_1)}\right\},\tag{30}$$

where $\omega_0=g\beta M_0$ is the frequency of the homogeneous ferromagnetic resonance.

3. SOLITON MOTION IN A NONUNIFORM MAGNETIC FIELD

We now turn to the more general problem of the dynamics of a soliton moving in a one-dimensional uniaxial ferromagnet placed in a nonuniform magnetic field $\mathbf{H}(x)$ that is parallel to the anisotropy axis. We assume that the function $H(x)$ varies in an arbitrary manner as a function of x but that the characteristic spatial scale over which $H(x)$ varies is much larger than the soliton width. Since the magnetic energy density of a uniaxial ferromagnet is independent of the phase φ, the energy E and the number N of magnons remain constants of motion. The dependence of the soliton energy on the three dynamical variables $\mathbf{P}$, X, and N in the adiabatic approximation is an obvious generalization of (23), i.e.,

$$E=E_0(P,N)+2\mu_0NH(X),\tag{31}$$

where $E_0(P,N)$ is still given by (14) and corresponds to the soliton energy in the absence of a magnetic field.

The time dependence of the quasimomentum P is given by an expression that is an obvious generalization of (17):

$$\frac{dP}{dt}=-2\mu_0N\,\frac{dH(X)}{dX}.\tag{32}$$

Equation (32) together with (31) and (14) solve the problem of the time dependence of P. However, it is more convenient to study the motion of the center of gravity directly by using the explicit form of single-soliton solutions of the Landau–Lifshitz equations.

In the reference frame attached to the center of gravity ($\xi=x-Vt$), the single-soliton solution of Eqs. (2) for a uniform magnetic field has the form[1]

$$\tan^2\frac{\theta(\xi)}{2}=\frac{2\kappa}{(\kappa_M-\kappa_m)\cosh(2\kappa\xi)+\kappa_m+\kappa_M-2\kappa},\tag{33}$$

where $\kappa=(E-2\mu_0NH)/2W_0l_0$ is the inverse of the soliton width, $\kappa_m=l_0^{-1}\tanh(N/N_1)$, $\kappa_M=1/\kappa_ml_0^2$, and the soliton velocity V and the angular precession frequency Ω are related to the parameter κ as follows:

$$V=V_ml_0\sqrt{(\kappa-\kappa_m)(\kappa_M-\kappa)},\tag{34}$$

$$\Omega=gH+gM_0\beta[1-(\kappa l_0)^2-(V/V_m)^2].\tag{35}$$

Since the parameter κ depends on the soliton's total energy and coordinate of the center of gravity, it is convenient to take it as the dynamical characteristic of the soliton. In particular, Eq. (34) shows that the possible movements of the soliton are limited to the range of values of κ for a fixed N: $\kappa_m\leqslant\kappa\leqslant\kappa_M$.

The motion of a soliton in a slowly varying magnetic field can be described in the adiabatic approximation by the same Eqs. (33)–(35) if we put $\xi=x-X(t)$ in them. Here the position of the center of gravity $X(t)$ of the soliton at time t is uniquely determined by the laws of conservation of the energy and the projection of the total magnetic moment on the z axis. In a nonuniform magnetic field, the parameters E and N, which enter into Eqs. (33)–(35), are still constants of motion, and the velocity V, the precession frequency Ω, and the inverse of the soliton width, κ, become slowly varying functions of time, and functionally the quantities V, Ω, and κ are determined by the running values of the magnetic field H at the soliton's center of gravity $X(t)$:

$$\kappa(X)=\frac{E-2\mu_0NH(X)}{2W_0l_0},$$

$$\Omega(X)=gH(X)+gM_0\times\left\{2-l_0(\kappa_M+\kappa_m)\,\frac{E-2\mu_0NH(X)}{2W_0}\right\},\tag{36}$$

$$V(X)=\frac{2g\mu_0N}{M_0}\{[H(X)-H_m][H_M-H(X)]\}^{1/2},\tag{37}$$

where $H_m=(E-2W_0l_0\kappa_M)(2\mu_0N)^{-1}$ and $H_M=(E-2W_0l_0\kappa_m)(2\mu_0N)^{-1}$ are the values of the magnetic field at the turning points x_m and x_M of the soliton (Fig. 1).

Integrating the equation $dX/dt=V(X)$ together with Eq. (37), we obtain the dependence of the coordinate X of the center of gravity on time t in the implicit form

$$t=\frac{N_1}{gl_0N}\,[F(X)-F(X(0))],\tag{38}$$

where

$$F(X)=\int\frac{dX}{\{[H_M-H(X)][H(X)-H_m]\}^{1/2}}.$$

If the magnetic field $H(x)$ reaches both characteristic values H_m and H_M for finite x, the soliton is in a state of bounded motion between the turning points x_m and x_M (Fig. 1a) with a period

$$T=\frac{N_1}{gl_0N}\,[F(x_M)-F(x_m)].\tag{39}$$

In a field with a constant magnetic-field gradient $\eta=dH/dx$, the soliton is in a state of harmonic oscillatory motion, where neither the frequency $\Omega_0=2gl_0(N/N_0)|\eta|$ nor the oscillation amplitude $\Delta X=|x_M-x_m|/2=(N_1\beta M_0/N|\eta|)/\sinh(2N/N_1)$ depends on the soliton energy E; they depend only on the number N of magnons. If

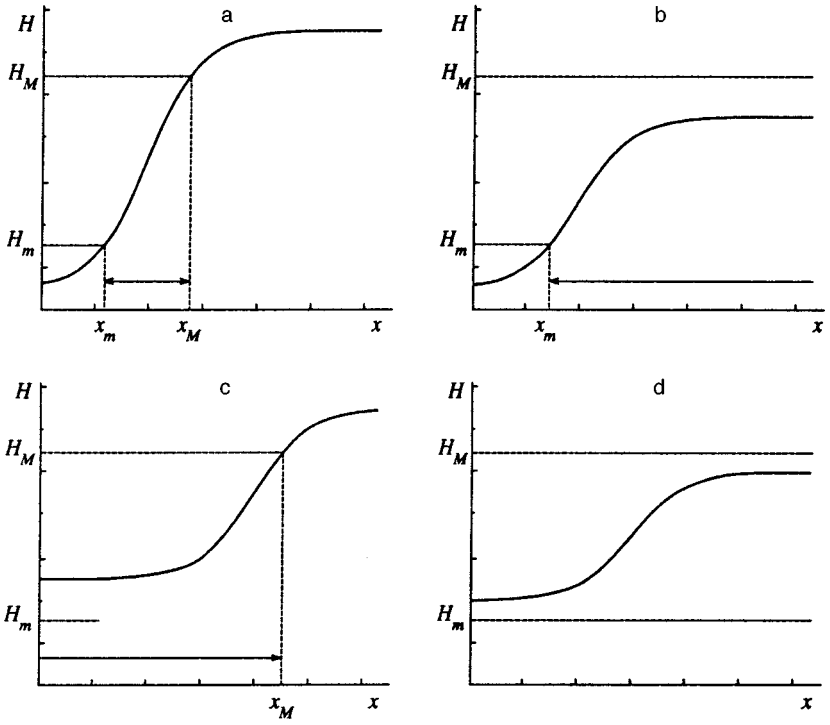


FIG. 1. Soliton motion in a nonuniform magnetic field: (a) bounded motion, (b) inbounded motion in the direction $x=\infty$, (c) unbounded motion in the direction $x=-\infty$, and (d) unbounded above-the-barrier motion.

$H(x)$ reaches only one value, H_m or H_M, the soliton is in unbounded motion, the reflection of the magnetic soliton from a ''magnetic potential barrier'' (Figs. 1b and 1c). If $H_m<H(x)<H_M$, the unbounded motion takes place above the barrier (Fig. 1d).

4. COMPUTER SIMULATION OF SOLITON OSCILLATIONS IN A NONUNIFORM FIELD

To verify our results we used a computer to numerically solve the one-dimensional Landau–Lifshitz equations

$$\frac{\partial \mathbf{M}}{\partial t}=g\mathbf{M}\times\left[\beta\mathbf{n}(\mathbf{M}\cdot\mathbf{n})-\alpha\frac{\partial^2\mathbf{M}}{\partial x^2}+\mathbf{H}(x)\right], \tag{40}$$

where the unit vector **n** is directed along the anisotropy axis (the z axis), and the nonuniformity of the magnetic field is characterized by a constant value of the field's gradient $\eta=dH/dx$. Equations (40) were solved by a standard fourth-order Runge–Kutta method with automatic selection of the timestep, and the spatial derivatives $\partial^2\mathbf{M}/\partial x^2$ were calculated via a five-point finite-difference approximation; the number of grip points in the variable x was $n=400$. At first glance it seems that Eqs. (40) are entirely useless, since in addition we must know the soliton's initial state in the nonuniform magnetic field, and this is unknown. Hence we employed the following method of ''preparing'' a soliton with fixed values of the parameters E and N in the nonuniform magnetic field. First we constructed an effective magnetic field $H(x)$ by joining a horizontal section $H(x)=\text{const}$ with a slanted section with the given gradient dH/dx (Fig. 2a). Then a soliton was placed in the horizontal section (curve *1*), for which Eqs. (33)–(35) were used. Finally the procedure of numerical solution of Eqs. (40) was initiated. As the soliton moves it enters the slanted section, where the field is nonuniform (curve *2*). At the moment when the soliton is entirely in the slanted section the horizontal section is replaced by a slanted one (Fig. 2b). Since the amplitude of magnetization oscillations in this section is infinitesimal, the soliton is ''unaware'' of such a substitution (curve *2′*).

Figure 3 depicts the results of computer simulation of the soliton motion with the following values of the parameters:

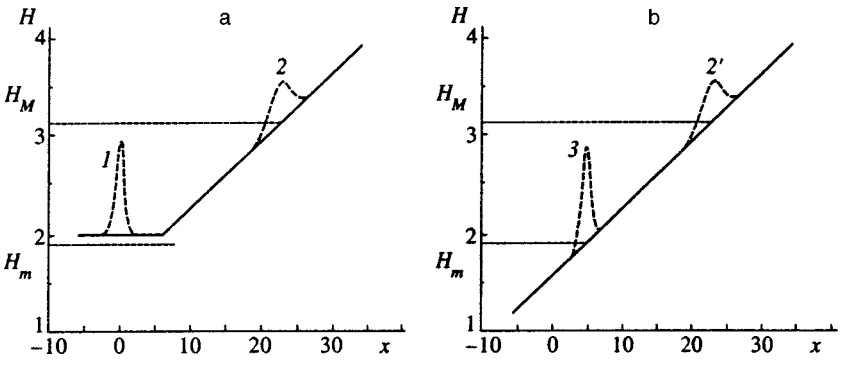


FIG. 2. ''Preparation'' of a soliton with given values of the parameters E and N: (a) the soliton is in an effective field, and (b) the soliton is in a field with a constant gradient.

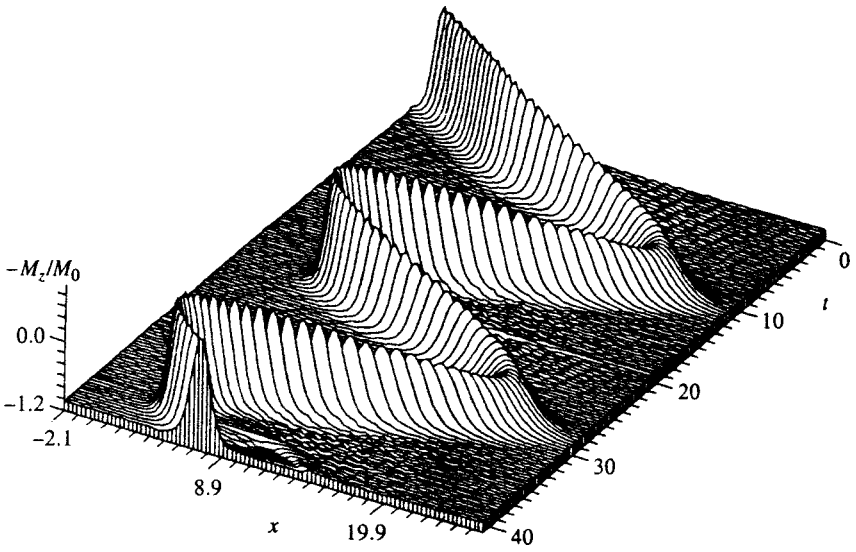


FIG. 3. Oscillatory motion of a soliton in a nonuniform magnetic field.

$$\frac{N}{N_1}=0.75, \quad \frac{E}{W_0}=6, \quad \frac{l_0|\eta|}{\beta M_0}=0.07.$$

The results of our calculations agree fairly well with those found from the formulas of the adiabatic theory. For instance, for the given values of the parameters, the theoretical values of the period and amplitude of the oscillations are $T=59.8/gM_0\beta$ and $\Delta X=8.95l_0$. The numerical calculations yield $T=60.4/gM_0\beta$ and $\Delta X=8.7l_0$, which are close to the theoretical values. However, more accurate calculations show that near the right turning point x_M the soliton emits a small-amplitude spin wave whose frequency corresponds to the frequency of magnetization precession in the soliton, Ω (Fig. 4). This phenomenon is related to the violation of adiabaticity and is described by higher-order corrections in the magnetic-field gradient.

The violation of the adiabatic approximation results from the interaction of the soliton and the spin waves, where the spin waves with the frequency Ω play the major role. The domain of existence of the waves, $x<x_s$, is bounded by the turning point x_s for these spin waves, where x_s can be found from the equation $gH(x_s)+gM_0\beta=\Omega(X)$.

Using the formulas in (36), we can calculate the distance between the soliton's center of gravity and the boundary of the spin-wave range:

$$|X-x_s|=\left[(1+\kappa_m^2 l_0^2)\,\frac{\kappa(X)}{\kappa_m}-1\right]\frac{\beta M_0}{|\eta|}. \tag{41}$$

We see that the distance is inversely proportional to the field gradient η and depends on X, the coordinate of the center of gravity.

According to (33), the amplitude of magnetization oscillations in the soliton falls off at large distance as $e^{-\kappa|\xi|}$. Thus, the interaction of the soliton and the spin waves is characterized by the small parameter $\exp(-\kappa|X-x_s|)$. Correspondingly, the adiabatic approximation holds if the distance from the center of gravity X to the spin-wave turning point x_s is much larger than the soliton size κ^{-1}:

$$|X-x_s|\kappa(X)\gg 1. \tag{42}$$

We note that the adiabatic approximation may be valid in one spatial region and invalid in another. According to (41), the distance $|X-x_s|$ reaches its minimum value $\beta M_0 l_0^2\kappa_m^2/|\eta|$ near the right turning point. At this point the adiabaticity condition (42) amounts to the requirement that the magnetic-field gradient be small:

$$\left|\frac{dH}{dx_M}\right|\ll\beta M_0 l_0^2\kappa_m^3. \tag{43}$$

Hence, the adiabatic theory that we developed to describe the soliton motion in a weakly nonuniform magnetic field is valid if

$$\frac{l_0}{\beta M_0}\left|\frac{dH}{dx_M}\right|\Big/\tanh^3\frac{N}{N_1}\ll 1. \tag{44}$$

As the magnetic-field gradient gets stronger, the adiabatic approximation is violated first of all near the turning point that corresponds to the maximum admissible value of the magnetic field (which in our case is the right turning point).

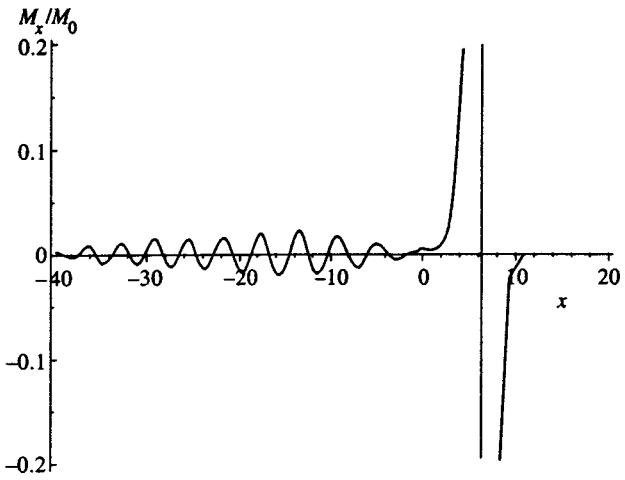


FIG. 4. Emission of a spin wave by a magnetic soliton.

For the values of the parameters used in the computer simulation of soliton oscillations, the left-hand side of (44) is $\varepsilon = 0.3$, which is the limit of the applicability of the adiabatic approximation. Near the left turning point the condition (44) is replaced by a less stringent condition,

$$\frac{l_0}{\beta M_0}\left|\frac{dH}{dx_m}\right|\tanh^3\frac{N}{N_1} \ll 1, \tag{45}$$

whose left-hand side in our example is $\varepsilon = 0.017$. Note that $\exp(-1/\varepsilon)$ is the small parameter in the adiabatic approximation.

Thus, a magnetic soliton placed in a nonuniform magnetic field can be in a state of periodic motion with a frequency Ω_0 of order 10^9 Hz.

*)E-mail: kosevich@ilt.kharkov.ua

[1] A. M. Kosevich, B. A. Ivanov, and A. S. Kovalev, *Nonlinear Magnetization Waves. Dynamic and Topological Solitons* [in Russian], Naukova Dumka, Kiev (1983).
[2] A. M. Kosevich, B. A. Ivanov, and A. S. Kovalev, Phys. Rep. **194**, 117 (1990).
[3] A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskiĭ, *Spin Waves*, Interscience, New York (1968).
[4] A. M. Kosevich, Physica D **119(1–2)**, 134 (1998).
[5] V. I. Karpman and E. Maslov, Zh. Éksp. Teor. Fiz. **73**, 537 (1977) [Sov. Phys. JETP **46**, 281 (1977)].
[6] D. J. Kaup and A. C. Newell, Proc. R. Soc. London, Ser. A **361**, 413 (1978).
[7] V. I. Karpman and E. Maslov, Phys. Fluids **25**, 1682 (1982).

Translated by Eugene Yankovsky

Structure of zero modes in a model of the discrete $(2+1)$-dimensional nonlinear Schrödinger equation

L. A. Abramyan and A. P. Protogenov*)

Institute of Applied Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia

V. A. Verbus

Institute of Microstructure Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia

(Submitted 16 February 1998)
Zh. Éksp. Teor. Fiz. **114**, 747–762 (August 1998)

The structure of the zero modes in a discrete $(2+1)$-dimensional model of the gauge-invariant nonlinear Schrödinger equation is studied. Including the compactification of the Chern–Simons gauge fields eliminates the difficulties with the continuous model [L. A. Abramyan and A. P. Protogenov, JETP Lett. **64**, 859 (1996); L. A. Abramyan, V. I. Berezhiani, and A. P. Protogenov, Phys. Rev. E **56**, 6026 (1997)] and leads to a prediction of the existence of a transition region characterized by a hierarchical sequence of collapses which are enumerated by the Chern–Simons coefficient. Using the zero modes in calculating the dependence of the critical power N on the Chern–Simons coefficient, we have found that the transition region lies in the interval $11.703 \leqslant N \leqslant 12.01$. © *1998 American Institute of Physics.* [S1063-7761(98)02908-4]

1. INTRODUCTION

The nonlinear Schrödinger equation (NLSE) is one of the basic models in the theory of nonlinear waves. The traditional domain of application of the NLSE is nonlinear optics,[2,3] where it describes the propagation of wave beams in dispersive nonlinear media. In this case, the time variable is replaced by the axial coordinate z along the beam in the $(2+1)$-dimensional NLSE. The NLSE also appears in studies of various nonlinear waves in hydrodynamics and plasma physics.[4] One of the most important applications in this case is the detailed description of collapsing field distributions both using the NLSE with a local cubic nonlinearity,[5,6] and in terms of the Zakharov equations.[7] In the case of a nonlinearity of opposite sign (repulsion), the NLSE is used, for example, to describe vortices in the Bose condensation problem,[8] and as a basis model[9] for a low-dimensional field theory.

Recently there has been increased interest in the NLSE for $(2+1)$-dimensional systems as a result of attempts to allow for the topological features of the manifold on which the field is defined in spatially two-dimensional systems. This can be done with the aid of a gauge field and by replacing the ordinary derivatives with covariant derivatives in the equations of motion. The gauge field satisfies its own equation of motion with a flux defined by the solutions of the NLSE and, as an auxiliary variable, describes an additional contribution to the nonlinearity of the standard NLSE. In the infrared limit the main contribution in the equations of motion for a gauge field in a $(2+1)$-dimensional system is the Chern–Simons term in the action for the system under consideration. For a certain relationship among the coupling constants, the contribution of the gauge field compensates the contribution of the nonlinearity leading to collapse in the Hamiltonian. In this case the soliton field distributions reported by Jackiw and Pi[10] are formed. This phenomenon is easily understood if we note that in $(2+1)$-dimensional systems the Chern–Simons term breaks the symmetry with respect to $\mathscr{P}$ and $\mathscr{T}$ inversion of the coordinates and time. A preferential orientation of a vector in a direction perpendicular to the plane can be perceived as a preferential direction of rotation in a plane, which leads to the appearance of an effective repulsion. In a situation where this repulsion compensates the attraction inherent in the NLSE, the Hamiltonian ends up having a lower bound, and its zero-point value corresponds to self-dual Chern–Simons solitons.[10] These field distributions are solutions of the duality equations.

A stream of papers in this area has been stimulated by the results of Jackiw *et al.*[10,12] and Hong *et al.*[11] Let us focus our attention on some of them. Barashenkov and Harin[13] performed a detailed analysis of the structures of the field configurations for a nonlinear function in the NLSE which describes a repulsion (in the absence of a Chern–Simons interaction) and includes a contribution from a nonzero vacuum average for the particle number density. A formulation of the initial problem by Bergé *et al.*[14] led them to the conclusion that, under the most general initial conditions, the solutions of the equations of motion for the NLSE with allowance for a Chern–Simons gauge field at a negative value of the Hamiltonian correspond to collapse. However, in that paper, neither the spatial structure of the collapsing mode nor its critical power (the number of particles in the mode) were analyzed. The integrability of this model was the subject of Ref. 15 The main result is that the system is not exactly integrable, except in two cases: the self-dual limit[10] and the situation where the $(2+1)$-dimensional equations can be reduced to a $(1+1)$-dimensional model through a replacement

of variables. An incidental, but no less important conclusion is that the solutions of the gauge-invariant NLSE have nonstationary singularities on several curves in a two-dimensional plane. A detailed study of topological defects in low-dimensional systems has always been, to a considerable extent, fundamental to understanding the dynamics of field distributions. A recent paper[16] was devoted to the problem of studying so-called semilocal topological defects in the extended Chern–Simons–Higgs model.

The main reason for the extremely specific behavior of the field distributions $\Psi(x,y,t)$ in spatially two-dimensional systems is that the space $\mathcal{M}$ in which the complex-valued functions are defined is multiply connected in this case. For this reason, the fundamental homotopic group $\pi_1(\mathcal{M})$, which determines the analytic properties of the function Ψ with respect to transformations of its arguments, does not coincide with the permutation group in $(2+1)$-dimensional systems, but with the braid group. There are several essentially equivalent ways of representing this circumstance in the methods used in the theory. One is the Lagrange approach, which incorporates the Chern–Simons action. In the long-wavelength approximation the Chern–Simons term encodes the existence and specific features of the spatially two-dimensional point singularities contained in the Bohm–Aharonov potentials of the gauge field. The long-range interaction, represented through the Chern–Simons gauge field, is usually referred to as a statistical or correlation interaction between different field configurations. The distribution of the field Ψ itself can then be assigned different forms, depending on the representation. There is the so-called anyon representation,[17,18] where the gauge field is explicitly excluded from the Hamiltonian for the model to ensure a picture of ''noninteracting'' (with the aid of a gauge field) field configurations $\Psi(x,y,t)$. In this case, however, the gauge field is included in the phase of the function $\Psi(x,y,t)$, which contains a section in the complex plane that ensures that this function is multivalued. The section describes a string stretching toward a point defect (a so-called nonlocal topological defect), which separates the sheets in the multisheet covering of the two-dimensional basis space. It is well known[17,18] that a representation in which the gauge field is explicitly present in a Lagrange model can have different forms, depending on the parity of the representation of the permutation group. The correlation effects, however, do not depend on the representation, and the anyon statistics take on the form of the dynamics of the long-range gauge field in this case. From this standpoint, an approach based on a representation with fractional statistics of the field configurations is identical to the dynamic approach, where we are interested in the form of field distributions that are subjected to the influence of a ''statistical'' gauge field.

Although the reason for the existence of the gauge interaction is exclusively topological (geometric) and is not related to quantum theory, this interaction has generally (except for the above mentioned papers) not been taken into account in studies of the classical dynamics of nonlinear models with a complex field in spatially two-dimensional systems. The topological features, of course, impose additional limitations on the quantization procedure in systems of this type.[18] An important common feature of both the quantum and classical theories is the use of a complex function $\Psi(x,y,t)$ with a phase distribution which takes the form of the phase dynamics in the classical region. The role of the Chern–Simons gauge interaction in this case is to account for the vortex part of the phase dynamics, which has usually been neglected in classical systems using a model of the $(2+1)$-dimensional NLSE.

In this paper our goal is to clarify the role and magnitude of the topological effects associated with the violation of chiral invariance in classical $(2+1)$-dimensional systems. From this standpoint, the NLSE can be regarded as a useful example for solving this problem. Thus, in this paper we continue the earlier studies[14,15,1] of the equations of motion in a $(2+1)$-dimensional model for the gauge-invariant NLSE. In Ref. 1, primary attention was devoted to studying the structure of the collapsing field distribution which is a solution of the $(2+1)$-dimensional NLSE in the continuum limit. In particular, numerical integration of the equation of motion yielded the dependences of the critical power and effective width of the zero mode on the coefficient k in front of the Chern–Simons term. We used the limit $k\to\infty$, where the interaction with the gauge field is negligibly small, as a test. In this case, the known values of the power and width were reproduced. At low k, however, this computational scheme led to a divergence in the critical power. In the present paper we show that including nonperturbative values of the gauge field makes it possible to move into the region of small k. Introducing a solely spatial lattice makes it possible to only partially allow for the contribution of gauge fields of finite amplitude. Thus, in order to solve the problem stated here, we include two conditions in the analysis. First, this model is placed on a two-dimensional lattice. We note that the stability of NLSE solitons on a lattice without a gauge field has been studied in detail recently.[19,20] Second, we include the discreteness of time. Here discrete evolution is a necessary condition. The motive for introducing a discrete time is the requirement for a unified description of the contribution from large amplitude spatial and temporal components of the gauge potential. This approach is supported by recent papers on discrete dynamics.[21]

If the phase of the field $\Psi(x,y,t)$ completely describes the longitudinal part in the gauge potential, then the evolution of the field configurations is determined only by the time dependence of the gauge field. In this case the conservation laws in the Chern–Simons system, viz., Gauss' law and the conservation of the number of particles, are equivalent to the equations of motion of an ideal fluid in vortex form. Manifestations of gauge coupling in classical systems with nontrivial topologies, including cases in two-dimensional hydrodynamics, are well known, in particular, in the description of swimming motions at low Reynolds numbers.[22] The new factor in this connection is that the justification for two-dimensional turbulence based on the Euler equations is, in this case, the dynamics of the Chern–Simons gauge field. In this sense, the gauge-invariant NLSE is a useful tool in hydrodynamics.[23]

Regarding the reason for the appearance of Chern–Simons fields in the two-dimensional turbulence problem,

we would like to note the following. It is well known that Chern–Simons action with suitable boundary conditions is a means for classifying conformal theories.[24] The machinery of conformal field theory can, in turn, be applied[25] to the study of two-dimensional turbulence. The observation mentioned above actually means that, in terms of this model, a connection between the dynamics of the Chern–Simons fields and two-dimensional turbulence can be established without using conformal field theory.

This paper is organized as follows. For completeness of the discussion and clarification of all details of the problem, in the second section we formulate the problem of and present results pertaining to solutions of the equations of motion of this model in the continuum limit. The third section is devoted to stating and solving the main problem of this paper: analyzing the structure of the fundamental modes and their contribution to the integrals of motion in the discrete variant of the model. In the fourth section results from a numerical analysis of the problem are presented. In the final, fifth, section, some open questions, as well as the areas of applicability of these results, are discussed.

2. EQUATIONS OF MOTION

In the continuum limit the density of the Lagrangian for the system under consideration has the form

$$\mathscr{L}=\frac{k}{2}\,\varepsilon^{\alpha\beta\gamma}A_\alpha\partial_\beta A_\gamma+i\Psi^*(\partial_t+iA_0)\Psi - \frac{1}{2}\,|(\nabla-i\mathbf{A})\Psi|^2+\frac{g}{2}\,|\Psi|^4. \tag{1}$$

For Eq. (1) we write the equations of motion

$$i\partial_t\Psi=-\frac{1}{2}\,(\nabla-i\mathbf{A})^2\Psi+A_0\Psi-g|\Psi|^2\Psi, \tag{2}$$

$$(\operatorname{curl}\mathbf{A})_\perp=-\frac{1}{k}\,|\Psi|^2, \tag{3}$$

$$\partial_t A_i+\partial_i A_0=-\frac{1}{k}\,\varepsilon_{ij}j_j\,. \tag{4}$$

Here g is the coupling constant, $\mathbf{j}=\operatorname{Im}\Psi^*(\nabla-i\mathbf{A})\Psi$ is the current density, and k is the Chern–Simons coefficient. The Hamiltonian for Eq. (1) is

$$H=\frac{1}{2}\int d^2r(|(\nabla-i\mathbf{A})\Psi|^2-g|\Psi|^4), \tag{5}$$

where the potential A_μ is an auxiliary variable expressed in terms of $|\Psi|^2$ as follows:

$$\mathbf{A}(\mathbf{r},t)=\frac{1}{k}\int d^2r'\mathbf{G}(\mathbf{r}-\mathbf{r}')\rho(\mathbf{r}',t), \tag{6}$$

$$A_0(\mathbf{r},t)=\frac{1}{k}\int d^2r'\mathbf{G}(\mathbf{r}-\mathbf{r}')\cdot\mathbf{j}(\mathbf{r}',t). \tag{7}$$

The Green's function $\mathbf{G}(\mathbf{r})$,

$$G_i(\mathbf{r})=\frac{1}{2\pi}\,\frac{\varepsilon_{ij}x_j}{r^2}, \tag{8}$$

satisfies the equation

$$\operatorname{curl}\mathbf{G}(\mathbf{r})=-\delta^2(\mathbf{r}), \tag{9}$$

so that A_μ is a solution of Eqs. (3) and (4). Since in the Hamiltonian formulation the potentials are uniquely represented by Eqs. (6) and (7), the gauge freedom

$$A_\mu\to A_\mu-\partial_\mu\varphi, \tag{10}$$

$$\Psi\to e^{i\varphi}\Psi \tag{11}$$

is fixed. This is achieved by choosing the Coulomb condition $\operatorname{div}\mathbf{A}=0$ supplemented by the boundary conditions

$$\lim_{r\to\infty} r^2A_i(\mathbf{r},t)=\frac{1}{2\pi k}\,\varepsilon_{ij}x_jN, \tag{12}$$

$$\lim_{r\to\infty} A_0(\mathbf{r},t)=0. \tag{13}$$

The choice of boundary condition (12) is related to the need to satisfy the integral representation (3) of Gauss' law for Chern–Simons dynamics:

$$\Phi=\int d^2r(\operatorname{curl}\mathbf{A})_\perp=-\frac{1}{k}\int d^2r|\Psi|^2. \tag{14}$$

Here the magnetic flux Φ and the number of particles

$$N=\int d^2r|\Psi|^2 \tag{15}$$

are conserved quantities that ensure the global constraint $\Phi=-N/k$, which has the significance of Gauss' law for the Chern–Simons system.

As a consequence of Eqs. (2)–(4) there is a continuity equation

$$\partial_t|\Psi|^2+\operatorname{div}\mathbf{j}=0, \tag{16}$$

which expresses the time independence of N.

We now proceed, for convenience, to dimensionless variables and coordinates using the substitutions

$$\Psi=|k|^{3/2}\rho e^{i\varphi},\quad A_0=-\frac{k^2}{2}\,w-\partial_t\varphi,$$

$$A_x=-ku+\partial_x\varphi,\quad A_y=-kv+\partial_y\varphi, \tag{17}$$

$$t\to-\frac{2}{k|k|}\,t,\quad x\to\frac{x}{|k|},\quad y\to\frac{y}{|k|}. \tag{18}$$

The equations of motion and the continuity equation, expressed in terms of the new real functions $\rho\equiv\rho(x,y,t)$, $u\equiv u(x,y,t)$, $v\equiv v(x,y,t)$, and $w\equiv w(x,y,t)$, have the form

$$\rho_{xx}+\rho_{yy}=-2C\rho^3+\rho(u^2+v^2-w), \tag{19}$$

$$u_y-v_x=-\rho^2, \tag{20}$$

$$u_t-w_x=-2v\rho^2, \tag{21}$$

$$v_t-w_y=2u\rho^2, \tag{22}$$

$$\rho_t^2=2[(u\rho^2)_x+(v\rho^2)_y] \tag{23}$$

with the parameter $C=g|k|$ and the notation $u_t=\partial_t u$, etc.

In the case of the standard NLSE

$$i\partial_t\Psi = -\nabla^2\Psi - |\Psi|^2\Psi, \tag{24}$$

after substituting $\Psi = \rho\ \exp[-i\varphi(x,y,t)]$, we obtain

$$\rho_{xx} + \rho_{yy} = -\rho^3 + \rho[(\varphi_x)^2 + (\varphi_y)^2 - \varphi_t], \tag{25}$$

$$(\rho^2)_t = 2[(\varphi_x\rho^2)_x + (\varphi_y\rho^2)_y]. \tag{26}$$

A comparison of Eqs. (25) and (26) with Eqs. (19) and (23) reveals the following differences. First, because of the gauge invariance in Eq. (19), the phase derivatives φ, which exist in Eq. (25), are absent. Their role is played by scalar and vector potentials. Thus, the evolution of the field $\rho(x,y,t)$ is determined by time derivatives of the functions $u(x,y,t)$ and $v(x,y,t)$ in Eqs. (21) and (22). The fields u and v are responsible for the transverse dynamics of the phase of the field Ψ, in contrast to Eq. (25). The longitudinal dynamics of the phase are described by the scalar potential $w(x,y,t)$, which replaces the function φ_t in Eq. (19). The function $w(x,y,t)$ plays the role of a Lagrange multiplier, permitting local allowance for the constraint (20), which is imposed by Gauss' law $\Phi = -N/k$. Second, the continuity equation (23), which replaces Eq. (26), is a direct consequence of Eqs. (20)–(22). It can be obtained by eliminating the scalar potential w from Eqs. (21) and (22), if Eq. (20) is used.

Let us consider the ansatz for the field $\Psi(x,y,t)$, which corresponds to the generalized lens transformation[6,14]

$$\Psi(\mathbf{r},t) = \frac{\Phi(\boldsymbol{\zeta},\tau)}{g(\tau)}\ \exp\left(-\frac{ib(\tau)\boldsymbol{\zeta}^2}{2} + i\lambda\tau\right). \tag{27}$$

Here $\boldsymbol{\zeta} = \mathbf{r}/g(\tau)$, $\tau = \int_0^t du[f(u)]^{-2}$, and $b(\tau) = -f_t f = -g_\tau g$. With this substitution, the gauge potential transforms[10] as follows:

$$\mathbf{A}(\mathbf{r},t) \to [g(\tau)]^{-1}\mathbf{A}(\boldsymbol{\zeta},\tau), \tag{28}$$

$$A_0(\mathbf{r},t) \to [g(\tau)]^{-2}[A_0(\boldsymbol{\zeta},\tau) - b(\tau)\boldsymbol{\zeta}\mathbf{A}(\boldsymbol{\zeta},\tau)] \tag{29}$$

with preservation of the coupling of (6) and (7), where the function $\rho = |\Phi|$. After these transformations, Eq. (2) changes form:

$$i\partial_\tau\Phi + (\beta\boldsymbol{\zeta}^2 - \lambda)\Phi = -\frac{1}{2}(\nabla - i\mathbf{A})^2\Phi + A_0\Phi - g|\Phi|^2\Phi, \tag{30}$$

since $\beta(\tau) = (b^2 + b_\tau)/2 = -f^3 f_{tt}/2$ in the case where $\varphi(x,y,t) \sim b(x^2+y^2)$ and $b(t) \neq t_0 - t$ is nonzero. However, if we are interested in the collapsing solutions[26] with $f^2(t) \sim (t_0 - t)/\ln[\ln(t_0 - t)]$, then the structure of the self-similar nonlinear core[14] is described by solutions of the following equation:

$$-\lambda\Phi = -\frac{1}{2}(\nabla - i\mathbf{A})^2\Phi + A_0\Phi - g|\Phi|^2\Phi. \tag{31}$$

Numerical integration was used in Ref. 1 to find a localized solution (with zero energy) of Eq. (31). Figure 1 shows the dependence of the critical power (the number N corresponding to the onset of collapse in the two-dimensional case) found using this solution on the parameter $C = g|k|$. Here we note that the normalizations of the function ρ^2 and, therefore, of the number of particles N in Fig. 1, differ from the normalizations in Eq. (17) of this paper and in Ref. 1 by a factor of $2C$, i.e., $\rho^2 \to 2C\rho^2$. These changes in notation stem from passage to a description using the quantities customarily employed to describe self-focusing phenomena. In the interval $1 < C < 2.83$ we were unable to do the calculations within a model with a continuous field distribution, since the convergence of the iterative calculation scheme broke down. This was one of our reasons for turning to discrete field dynamics. The formal reason for the divergence lay in a change in the sign of the right-hand side of Eq. (19) when $C \simeq 1$.

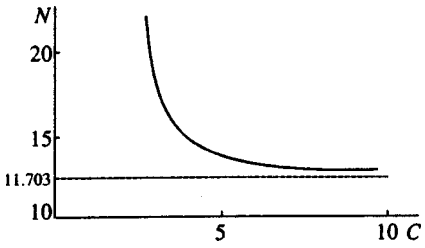


FIG. 1. Number of particles $N = \int d^2r\ \rho^2$ as a function of the parameter $C = g|k|$ for the gauge-invariant NLSE in the continuum limit.

For a fixed value of C in the range $C \geqslant 2.83$, we always have $N(A_\mu \neq 0) > N(A_\mu = 0)$, as is to be expected, since the Chern–Simons gauge field describes an effective repulsion. This result was obtained using a function with a Gaussian decay law (as a seed function) in the calculations. Note that the minimum possible value, $C = 1$ (i.e., to the discrete value $k = 1$ when $g = 1$), corresponds to the self-dual limit.[10] In this case, the function ρ exhibits power-law decay, and for axially symmetric fields $N = 4\pi = 12.56$. The classical region for this theory corresponds to the limit $k \to \infty$, where the gauge field splits off from the field $\Psi(x,y,t)$ [see Eq. (3)], and the critical self-focusing power is $N = 11.703$.

3. DISCRETE DYNAMICS

The reason for the difficulties in calculating N noted at the end of the previous section lies in the following. As the parameter $C = g|k|$ of the problem is reduced, the contribution from the nonlinearity induced by the gauge field, which is effectively proportional to ρ^5, becomes comparable to the contribution from the nonlinear term of the NLSE, which is proportional to $2C\rho^3$ [see Eq. (19)]. Since the iterative calculation scheme of Ref. 1 used a stabilizing factor that included both terms to some extent (which is determined by comparing the degrees of homogeneity of the functions), this stabilizing factor did not tend to unity at small C, where the terms became similar in magnitude.

In order to solve the problem for $C \simeq 1$, we point out the following circumstance. Equation (2) was written in the long-wavelength limit corresponding to gauge fields of small amplitude. An approach for finite values of the gauge fields is known. It is a formulation of the theory on a lattice using the derivatives

$$\Delta_\mu^+\rho(\mathbf{r}) \equiv \exp[iA_\mu(\mathbf{r})]\rho(\mathbf{r} + \mathbf{e}_\mu) - \rho(\mathbf{r}), \tag{32}$$

$$\Delta_\mu^-\rho(\mathbf{r}) \equiv \rho(\mathbf{r}) - \exp[-iA_\mu(\mathbf{r})]\rho(\mathbf{r} - \mathbf{e}_\mu) \tag{33}$$

and the Laplacian

$$\Delta(A)\rho_{m,n}\equiv\Delta^{+}\Delta^{-}\rho_{m,n}=\exp(iA_{m,\hat{n}})\rho_{m,n+1}+\exp(-iA_{m,\check{n}})\rho_{m,n-1}+\exp(iA_{\hat{m},n})\rho_{m+1,n}+\exp(-iA_{\check{m},n})\rho_{m-1,n}-4\rho_{m,n}. \quad (34)$$

Here and in the following we adhere to the following notation. The site coordinates $\mathbf{r}=(m,n)\in\mathbb{Z}^2$ comprise a discrete spatial variable, and the subscript μ indicates the direction of the unit vector $\mathbf{e}_\mu$ on the lattice and the component $A_\mu(\mathbf{r})$ of the gauge field in Eqs. (32) and (33). For convenience in the notation, we have returned to $\mathbf{r}$, keeping in mind that Eq. (30) in the variables $\boldsymbol{\zeta}$ for describing the core structure (when $\beta\ll 1$) and Eq. (3) in the variables $\mathbf{r}$ are the same. Several cells of the direct and dual lattices are shown in Fig. 2. In accordance with the rules of the gauge field theory on a lattice, we shall assume that the phase of the field $\Psi(\mathbf{r})$ is defined at sites $A,B,\ldots$ of the direct lattice, that the gauge field is defined at the links AB of the lattice, and that the curl of the field $A_\mu(\mathbf{r})$ and the density ρ^2 are defined at the sites $a,b,\ldots$ of the dual lattice. The notation $A_{m,\hat{n}}$ means that the field component $A_y(m,n)$ is defined on the link with end coordinates (m,n) and $(m,n+1)$. Accordingly, $A_{m,\check{n}}$ corresponds to defining the field $A_y(m,n)$ on the link with end coordinates (m,n) and $(m,n-1)$.

The same reason that made us have to formulate the theory on a spatial lattice, taking small values of the vector potential into account, dictates a need to add discrete time to the analysis. Here the form in which the time component of the gauge potential is represented in the equation of motion must be equivalent to that in which the spatial components of the gauge potential are represented. In other words, the Polyakov and Wilson indices (homologs) must be represented on the same footing. The fact that motion on a spatial lattice dictates the need for discrete evolution was noted for a hyperbolic operator in Ref. 21.

The above requirements of the theory correspond to the following replacement:

$$\left(i\frac{\partial}{\partial t}-A_0\right)\Psi\to\frac{i}{2}\left[\exp\left(\frac{\partial}{\partial t}+iA_0\right)-\exp\left(-\frac{\partial}{\partial t}-iA_0\right)\right]\Psi=-\Psi\sin(A_0+1). \quad (35)$$

The last equality in Eq. (35) is valid for the stationary states $\Psi(\mathbf{r},t)=\rho_{m,n}\exp(i\lambda t)$ to which we restrict ourselves in this paper. Without loss of generality, we shall assume that $\lambda=1$ (for further details, see the next section).

The equations of motion in the model of a discrete gauge-invariant NLSE with allowance for Eqs. (34) and (35) have the form

$$\exp(iA_{m,\hat{n}})\rho_{m,n+1}+\exp(-iA_{m,\check{n}})\rho_{m,n-1}+\exp(iA_{\hat{m},n})\rho_{m+1,n}+\exp(-iA_{\check{m},n})\rho_{m-1,n}-4\rho_{m,n}=-2C\rho_{m,n}^3-\rho_{m,n}\sin(w_{m,n}-1). \quad (36)$$

Here

$$A_{m,\hat{n}}=-\sum_{m',n'}\Delta_1 G(m-m',n-n')\rho^2_{m',n'}, \quad (37)$$

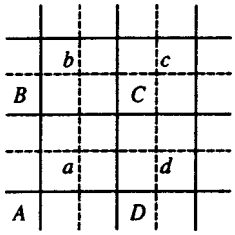


FIG. 2. Fragment of the direct and dual lattices with sites at the points $A,B,\ldots$ and $a,b,\ldots$, respectively.

$$w_{m,n}=\sum_{\mathbf{r}'}\left[(\Delta_2 G(\mathbf{r}-\mathbf{r}'))(\rho^2_{m',n'}+\rho^2_{m',n'+1})A_{m',\hat{n}'}-(\Delta_1 G(\mathbf{r}-\mathbf{r}'))(\rho^2_{m',n'}+\rho^2_{m'+1,n'})A_{\hat{m}',n'}\right] \quad (38)$$

and

$$\Delta_1 f(\mathbf{r})\equiv f(\mathbf{r}+\mathbf{e}_1)-f(\mathbf{r}).$$

Equations (37) and (38) are the discrete analogs of Eqs. (6) and (7). The sum over $\mathbf{r}'$ in Eqs. (37) and (38) corresponds to summing over all sites of the dual lattice. Here we use the convention that the coordinate of a site in the dual lattice, where the density ρ^2 is defined (point a in Fig. 2) corresponds to a certain (point A in Fig. 2) site in the direct lattice, where the starting point of the link on which the gauge field is determined lies. The Green's function on the lattice, which appears in Eqs. (37) and (38), has the form

$$G(m-m',n-n')=\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\frac{\exp i\{s(m-m')+t(n-n')\}}{4-2\cos s-2\cos t}\,ds\,dt. \quad (39)$$

For numerical solution of the problem we rewrite Eq. (36) in the following form:

$$\rho_{m,n+1}+\rho_{m,n-1}+\rho_{m+1,n}+\rho_{m-1,n}-4\rho_{m,n}=2C\rho^3_{m,n}+\rho_{m,n}[4-2\cos u_{m,n}-2\cos v_{m,n}-\sin(w_{m,n}-1)]. \quad (40)$$

In writing down the nonlinear terms on the right-hand side of this equation that are associated with the spatial components of the gauge potential, to simplify the numerical calculations we have neglected the difference between the functions in neighboring sites and links. We emphasize once again, in order to avoid misunderstandings, that we are introducing the model (40), which differs from Eq. (19), into the discussion in order to proceed to small values of the parameter C. For large C they coincide in the continuum limit.

The form of Eq. (40) satisfies the requirements formulated at the beginning of this section. In fact, trigonometric expressions of the type $\cos u$ and $\sin(w-1)$ reflect the impossibility of a large contribution from large amplitude gauge fields. In other words, regularization of the theory owing to the spatiotemporal lattice naturally leads to its compactification.

Some results of numerically integrating this equation, which give an idea of the structure of the zero mode and the

form of the auxiliary fields $A_\mu(m,n)$, are presented in the next section. The field configurations $\rho(\mathbf{r})$ found, which are used to calculate the number of particles N as a function of the parameter C, make it possible to clarify the details of the intermediate power self-focusing regime at small C.

4. NUMERICAL CALCULATIONS

For numerical analysis of the solutions of the equation of motion (40) we use the Petviashvili stabilizing factor method.[27] The iteration scheme for Eq. (40) has the form

$$\rho_{l+1}=M_l F^{-1}(G(p)F(-2C\rho_l^3+j\rho_l\{4-2\cos u - 2\cos v-[1+\sin(w-1)]\}_l)), \tag{41}$$

$$M_l=\left[\int d^2p(F\rho_l)^2\right]^\alpha\left[\int d^2pG(p)F\rho_l F(-2C\rho_l^3+j\rho_l \times\{4-2\cos u-2\cos v-[1+\sin(w-1)]\}_l)\right]^{-\alpha}. \tag{42}$$

Here $\rho_l\equiv(\rho_{m,n})_l$, the index l denotes the iteration number, F (F^{-1}) is the Fourier (inverse Fourier) transformation operator, and $G(p)=-(4-2\cos p_x-2\cos p_y+1)^{-1}$. Here $j=1$ or $j=0$, depending on whether we include or neglect the nonlinear contribution from the gauge fields in the structure of the zero modes.

The exponent α in the stabilizing factor M_l should be chosen from the requirement that $M_l\to 1$ as $l\to\infty$. In the case of homogeneous functions and without the nonlinear term $\rho(u^2+v^2-w)$, it follows from a comparison of the degrees of homogeneity of the terms on the left- and right-hand sides of Eq. (19) that $\alpha=3/2$. In small gauge fields, where the nonlinearity in Eq. (40) has a polynomial dependence of the type $-2C\rho^3=b\rho^5$ [since both ρw and $\rho(u^2+v^2)$ are proportional to ρ^5], the exponent α must lie within the interval $5/4\leqslant\alpha\leqslant 3/2$ for convergence of the iteration scheme. Although the nonlinear functions we are using now are inhomogeneous, we used the exponent $\alpha=3/2$ in the present numerical calculations; this yields a rapid approach to a value of $M_n=1$ for the stabilizing factor. For the initial field configurations we used distributions of the form $\rho(m,n)=(\gamma/\pi)\exp\{-\gamma(m^2+n^2)\}$ with $\gamma=2$. As opposed to the continuum limit,[1] in the calculations on a lattice we did not encounter difficulties associated with divergence of the expressions used, since our cutoff radius equalled the lattice spacing.

The modeling was done on a square lattice with a spacing of unity and with maximum linear dimensions of $L_x=L_y=20$. As a test, the solution of the equation of motion (40) with $A_\mu=0$ ($j=0$) in a $C=1/2$ normalization was used; this yielded the well-known value $N=11.703$. In addition, we compared the field distributions in the discrete and continuous cases at $C\geqslant 5$, where they practically coincided.

Figure 3 shows the configurations of the fields ρ, u, and w for the typical value $C=3$. Some idea of the function $v(m,n)$ can be obtained using the formula $v(m,n)=-u(n,m)$.

Using the zero mode $\rho_{m,n}$, we calculated the dependence of the mean square of the effective width $\langle R^2\rangle=N^{-1}\Sigma_{m,n}(m^2+n^2)/\rho_{m,n}^2$ and of the critical power N on C. The calculated values of $\langle R^2\rangle$ are listed in Table I. A plot of $N(C)$ for $C\geqslant 1$ is shown in Fig. 4.

5. DISCUSSION

An idea of the magnitude of the topological Chern–Simons effects can be obtained from Fig. 4. We see that allowing for compactification of the theory leads to a radical change in the critical power for self focusing phenomena. Correct allowance for the contribution of large-amplitude gauge fields led to a reduction in N for all C and made it possible to calculate the critical power for small C. In the theory which neglects the contribution of the gauge fields, there is a sharp boundary at $N=11.703$. Now we can see that the critical powers N corresponding to zero modes with different values of the Chern–Simons parameter k lie within the interval $11.703\leqslant N\leqslant 12.01$. We estimate the accuracy of our calculations as a few percent.

We note that, if we allow for the discrete character of the motion in time using the function $\sin(w-1)$, while retaining the nonlinear function u^2+v^2 (i.e., while disregarding the possibility of constraining the large-amplitude spatial components of the gauge potential through compactification of the theory), then we obtain the following result: up to $C=3$ we have $N=11.7$, which corresponds to the value without a gauge field. [Calculating the functions $u(m,n)$ and $v(m,n)$ shows that they are indeed small: $\max\{u_{m,n},v_{m,n}\}\leqslant 0.1$.] And only for small $C\leqslant 3$ is N observed to fall to values $N<11.703$.

The natural conclusion to be reached in this case is that allowance for discrete evolution should be accompanied by a quite definite deformation in the form of the nonlinear terms. An observation of this sort can be found in some recently published papers.[21,28] From this point of view, including the function $4=2\cos u-2\cos v$ is a deformation of the u^2+v^2 nonlinearity and is a natural partner of the discrete evolution expressed through the function $\sin(w-1)$. The difference in the trigonometric functions is related to the fact that, as opposed to Refs. 21 and 28, we are considering an equation of motion with a parabolic operator.

In going from Eq. (36) to Eq. (40) to implement a numerical scheme based on the Petviashvili stabilizing factor method, we assumed that $\rho_{m,n\pm 1}=\rho_{m\pm 1,n}=\rho_{m,n}$, $A_{m,\hat{n}\pm 1}=A_{m,\hat{n}}$, and $A_{\check{m}\pm 1,n}=A_{\check{m},n}$ in the nonlinear terms. We found indirect confirmation of this hypothesis in calculations where the main nonlinearity $-2C\rho_{m,n}^3$ was varied following the suggestions of others.[29,21]

The spatial distributions of the fields $\rho(m,n)$, $u(m,n)$, and $w(m,n)$ for $C=3$ are shown in Fig. 3. A reduction in the amplitude of $\rho(m,n)$ compared to the continuous case can be seen there. A rapid isotropization of the initially anisotropic Gaussian distribution could be seen in the course of calculating $\rho(m,n)$. Thus, in the subsequent numerical calculations, an isotropic field distribution $\rho(m,n)=(\gamma/\pi)\times\exp\{-\gamma(m^2+n^2)\}$ with $\gamma=2$ was used as a seed function. The results of calculating the mean square of the effective

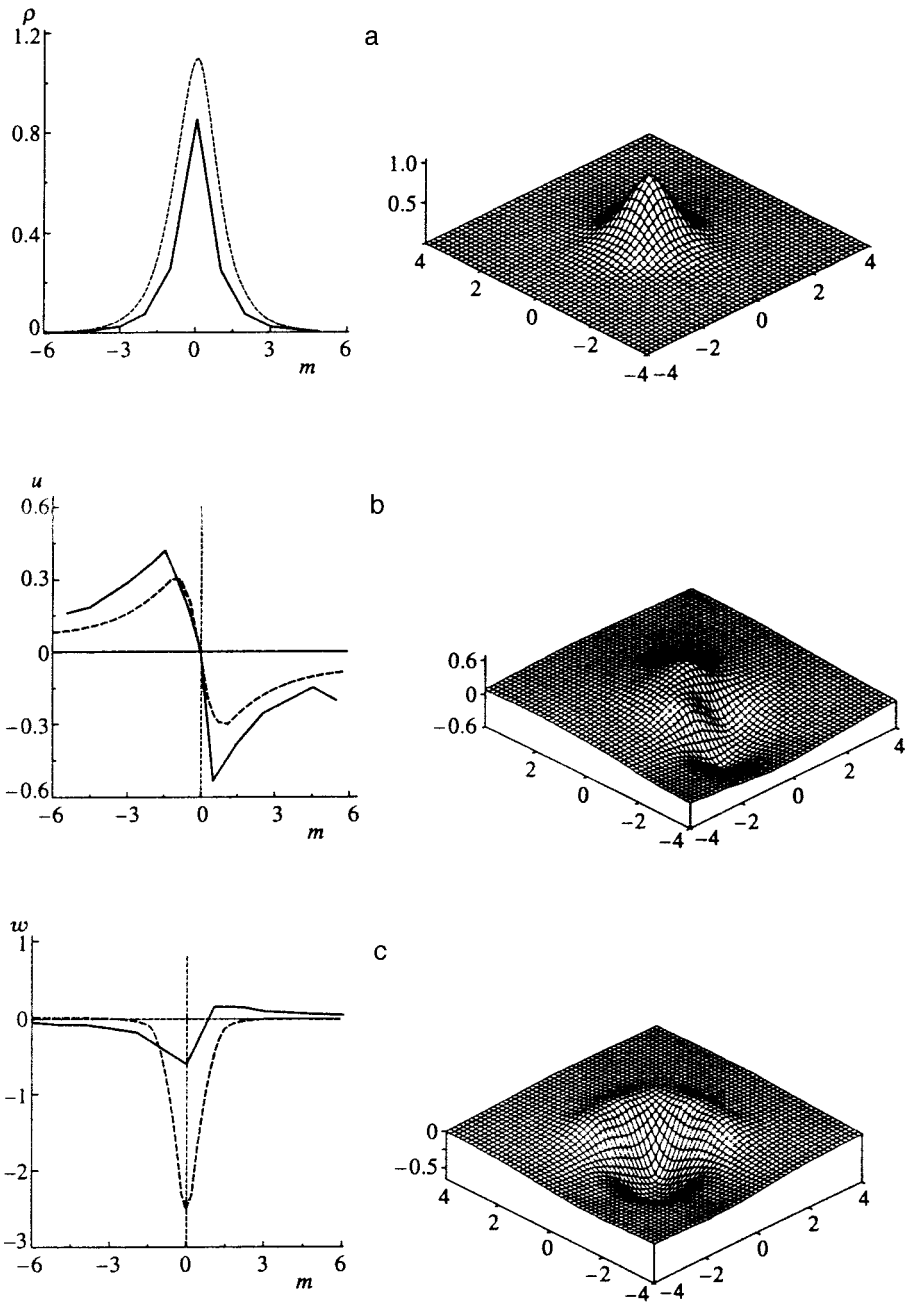


FIG. 3. Plots of $\rho(m)$, $u(m)$, and $w(m)$ for $C=3$ (left-hand sides of in Figs. a, b, and c, respectively) and the surfaces $\rho(m,n)$, $v(m,n)$, and $w(m,n)$ to the right of them. The functions in the continuum limit are indicated by dashed lines. The values of u and w for the discrete case in Figs. b and c are magnified by a factor of 2π.

width $\langle R^2\rangle$ of the function $\rho_{m,n}$ and its dependence on C are shown in Table I. Unlike the continuum model (19), here we observed a transition to an asymmetric distribution of w_m along the x axis. This phenomenon occurred only at small $C\leqslant 4$, where w_m becomes similar to $-u_m$.

In comparing these results, it should be kept in mind that the relationships between the amplitude function ρ, the parameter C, and the spatial scale length L for arbitrary λ (which are denoted by a tilde), on one hand, and the same quantities for $\lambda=1$ (which are denoted by a bar), on the other, are given by

$$\tilde{\rho}^2=\sqrt{\lambda}\overline{\rho}^2, \quad \tilde{C}=\sqrt{\lambda}\overline{C}, \quad \tilde{L}^2=\lambda^{-1}\overline{L}^2.$$

Therefore, the following chain of equalities holds:

$$\tilde{N}_{j\neq 0}=\overline{N}_{j\neq 0}/\sqrt{\lambda}=\overline{N}_{j=0}/2\sqrt{\lambda}\overline{C}=\tilde{N}_{j=0}/2\tilde{C}.$$

These equalities are useful for analyzing the contribution of the various terms on the right-hand sides of Eqs. (19) and (40) to the nonlinearity if $\lambda\neq 1$.

An idea of a few new possibilities which follow from these calculations can also be obtained from Fig. 4. They are related to the possible values of the coefficient k. Let $g=1$. In the case of an Abelian gauge field, the integer character of

TABLE I. Mean square of the width of the distribution of the field ρ for different values of the parameter C.

C	2	2.5	3	5
$\langle R^2\rangle$	2.57	2.60	2.64	2.72

the Chern–Simons coefficient k is an open question currently under intense discussion in the literature. In the non-Abelian case, where k is an integer or fraction (equal to the ratio of two mutually prime numbers), Abelian gauge fields can be regarded as added representations corresponding to the center of a non-Abelian group. Thus, we come upon the interesting possibility of comparing (with the aid of Fig. 4) discrete values of the critical power to k and thereby classifying the zero modes in terms of the values of this coefficient.

The usual treatment of Langmuir turbulence based on solving the NLSE relies on models of the appearance of a cascade of self-similar collapsing solutions. Now we can impart the following content to this picture. It is evident from Fig. 4 that the spectrum of critical powers becomes narrower with increasing k. This leads to a hierarchical picture of collapse, in which zero modes with different values of k are involved, and, as Eq. (18) implies, different spatial scales and characteristic times.

The coefficient k has the meaning of the number of points of interlocking of the world lines describing the evolution of the point singularities in a two-dimensional space. The small values of $C=2,3$ for $g=1$, where we observed significant differences from the classical results, correspond to singly $(k=2)$ and doubly $(k=3)$ interlocked world lines. We emphasize once again that these comments apply to the space of ground states that are degenerate in k. A description of the complete dynamics of the field configurations goes beyond the scope of this paper, but even now we can conclude that the discrete evolution of the zero modes we have considered will be a necessary element of a complete field description.

A fragment of this complete picture in the continuum model has been discussed previously[1] under the condition $\operatorname{div}\mathbf{A}=-u_x-v_y+\Delta\varphi=0$ when the phase φ satisfies the equation $\Delta\varphi=0$. In this case the phase $\varphi(x,y,t)$ is the linear function $\varphi=\alpha x+\beta y$, which, from the standpoint of the NLSE, corresponds (locally for small t) to a constant propagation direction of the rays specified by the vector $\mathbf{n}=(\alpha,\beta)$. In the case of self-focusing, this is not true, and the phase $\varphi(x,y,t)$, which does not obey the equation $\Delta\varphi=0$, creates an additional ''longitudinal'' contribution to the potentials $u(x,y,t)$ and $v(x,y,t)$ [see Eqs. (17), (27), and (30)]. The question of which of these possibilities is realized is related to the boundary conditions for the problem.

If the fields u and v do not contain a potential part, then the continuity equation (23) and Gauss' law (20) are equivalent to the equation of motion of an ideal Euler fluid in vortex form. For the stream function $a(x,y)$ this equation has the form

$$k\frac{\partial}{\partial t'}\Delta a+\frac{D(\Delta a,a)}{D(x,y)}=0.$$

Here Δ is the Laplacian and $D(\Delta a,a)/D(x,y)$ is the Jacobian. The time $t'=kt$ in this equation is reduced to dimensionless form with the aid of the coefficient k by analogy with the spatial coordinates x and y [see Eq. (18)]. In the limit $k\to\infty$ or, equivalently, $\Delta t'\to 0$, the Jacobian equals zero. This means that there is a functional relationship between Δa and a. Thus, in the case of small values of k, we necessarily encounter a nonstatic situation.

It can be presented in a graphic form by examining the evolution of the interlocked contours. The interlocking of closed lines of flow in the context of the present paper reflects the entanglement of the world lines of the point singularities with the formation of interlocking after projection of the world lines onto a two-dimensional space. Stochastization near the points of interlocking of the contours within a formulation of the two-dimensional hydrodynamics of an ideal liquid in terms of contour variables[30,31] has been observed.[32] This kind of random behavior has a universal character. It is characterized by arbitrariness in the location of the point of interlocking of the world lines along the time axis and, therefore (which is equivalent), to random positioning in the plane of the site where the closed contours are interlocked. The number of points of interlocking k is a hidden parameter, which does not appear explicitly in the Euler equation. This means that allowance for the specifics of spatially two-dimensional systems that are reflected in the Chern–Simons gauge fields can lead to the traditional picture of turbulence associated with the Euler equations. The existence of a close analogy between states with a constant flux in turbulence and the Chern–Simons anomalies expressed in Eq. (20) has been used to describe turbulence spectra elsewhere.[25]

One medium in which this mechanism for turbulence may operate is an optical medium with a random inhomogeneous distribution of reflecting surfaces in a channel. On being reflected from the surfaces, the wave fronts acquire random propagation directions, in which nonlinear phase shifts cannot develop.

We have, thus, studied the effect of the Chern–Simons gauge field, which reflects the specific features of the dimensionality of our problem, on the structure of the zero modes in a model of the discrete gauge-invariant NLSE. We have shown that correct allowance for large-amplitude gauge fields is important for $1<k<3$. We have also discovered the existence of a transition regime for self-focusing that is characterized by a finite range $11.703\leqslant N\leqslant 12.01$ of critical powers.

We would like to thank S. N. Vlasov, A. V. Gaponov-Grekhov, E. A. Kuznetsov, A. G. Litvak, V. A. Mironov, V. I. Talanov, G. M. Fraĭman, and A. D. Yunakovskiĭ for many stimulating discussions and useful comments. The numerical calculations were done on a DG III/ESPRIT work station provided by the Commission of the European Union, project CTIAC 21042. This work was partially supported by the Russian Fund for Fundamental Research (Grants No. 96-02-19272 (V.A.V.) and 98-02-16237 (L.A.A., V.A.V., and A.P.P.).

*)E-mail: alprot@appl.sci-nnov.ru

[1] L. A. Abramyan and A. P. Protogenov, JETP Lett. **64**, 859 (1996); L. A. Abramyan, V. I. Berezhiani, and A. P. Protogenov, Phys. Rev. E **56**, 6026 (1997).
[2] V. I. Talanov, JETP Lett. **2**, 138 (1965).

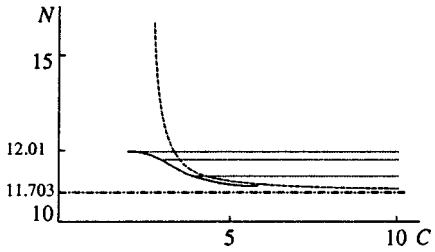


FIG. 4. Dependence of the number of particles N on the parameter C in the discrete model. The function $N(C)$ obtained in the continuum limit is shown by a dashed curve. Also shown is N corresponding to discrete k for $g=1$ (dotted curve). The dot-dashed curve corresponds to $k\to\infty$.

[3] S. N. Vlasov, V. A. Petrishchev, and V. I. Talanov, Izv. vuzov. Radiofizika **14**, 1453 (1971).
[4] *Singularities in Fluids, Plasmas, and Optics, NATO ASI Ser., Vol. C404*, R. E. Caflisch and G. C. Papanicolaou (eds.), Kluwer Acad. Publ., Dordrecht-Boston-London (1993); A. G. Litvak, ''Dynamic nonlinear electromagnetic phenomena in plasmas,'' in *Reviews of Plasma Physics, Vol. 10*, M. A. Leontovich (ed.), Consultants Bureau, New York (1986).
[5] A. G. Litvak, V. A. Mironov, and A. M. Sergeev, Phys. Scr. **T30**, 57 (1990).
[6] J. J. Rasmussen and K. Rypdal, Phys. Scr. **33**, 481 (1986).
[7] V. E. Zakharov, Zh. Éksp. Teor. Fiz. **62**, 1745 (1972) [Sov. Phys. JETP **35**, 908 (1972)].
[8] E. M. Lifshitz and L. P. Pitaevskii, *Statistical Physics* [in Russian], Part 2, Nauka, Moscow (1978), p. 145.
[9] E. A. Kuznetsov and S. K. Turitsin, Zh. Éksp. Teor. Fiz. **94**(8), 119 (1988) [Sov. Phys. JETP **67**, 1583 (1988)].
[10] R. Jackiw and S. Y. Pi, Phys. Rev. Lett. **64**, 2969 (1990); Phys. Rev. Lett. **66**, 2682 (1991); Phys. Rev. D **42**, 3500 (1990); Prog. Theor. Phys. Suppl. **107**, 1 (1992).
[11] J. Hong, Y. Kim, and P. Y. Pac, Phys. Rev. Lett. **64**, 2230 (1990).
[12] R. Jackiw and E. Weinberg, Phys. Rev. Lett. **64**, 2234 (1990); R. Jackiw, K. Lee, and E. Weinberg, Phys. Rev. D **42**, 3488 (1990).
[13] I. V. Barashenkov and A. O. Harin, Phys. Rev. Lett. **72**, 1575 (1994); Phys. Rev. D **52**, 2471 (1995).
[14] L. Bergé, A. de Bouard, and J. C. Saut, Phys. Rev. Lett. **74**, 3907 (1995).
[15] M. Knecht, R. Pasquier, and J. Y. Pasquier, J. Math. Phys. **36**, 4181 (1995).
[16] W. G. Fuertes and J. M. Guilarte, J. Math. Phys. **37**, 554 (1996).
[17] F. Wilczek, ed., *Fractional Statistics and Anyon Superconductivity*, World Scientific, Singapore (1990).
[18] A. P. Protogenov, Usp. Fiz. Nauk **162**, 1 (1992).
[19] E. W. Laedke, K. H. Spatschek, V. K. Mezentsev, S. L. Musher, I. V. Ryzhenkova, and S. K. Turitsyn, JETP Lett. **62**, 677 (1995).
[20] E. W. Laedke, K. H. Spatschek, and S. K. Turitsyn, Phys. Rev. Lett. **73**, 1055 (1994).
[21] L. D. Faddeev, *How Algebraic Bethe Ansatz Works for Integrable Model*, Les-Houches Lectures, hep-th/9605187.
[22] A. Shapere and F. Wilczek, J. Fluid Mech. **198**, 557 (1989).
[23] A. Shvartsburg, in *Nonlinear Electrodynamics* [Russian translation], P. L. E. Uslenfi, ed., Mir, Moscow (1980), p. 107.
[24] G. Moore and N. Seiberg, Phys. Lett. B **212**, 451 (1988); **220**, 422 (1989).
[25] A. M. Polyakov, Nucl. Phys. B **396**, 367 (1993).
[26] G. M. Fraĭman, Zh. Éksp. Teor. Fiz. **88**, 390 (1985) [Sov. Phys. JETP **61**, 228 (1985)].
[27] V. I. Petviashvili, Fiz. Plazmy **2**, 469 (1976) [Sov. J. Plasma Phys. **2**, 257 (1976)].
[28] V. Bazhanov, A. Bobenko, and N. Reshetikhin, Commun. Math. Phys. **175**, 377 (1996).
[29] M. J. Ablowitz and J. F. Ladik, Stud. Appl. Math. **55**, 213 (1976).
[30] N. J. Zabusky, M. H. Hughes, and K. V. Roberts, J. Comput. Phys. **30**, 96 (1979).
[31] A. A. Migdal, in *Nonlinear Waves. Structures and Bifurcations* [in Russian], A. V. Gaponov-Grekhov and M. I. Rabinovich, eds., Nauka, Moscow (1987).
[32] M. E. Aginshteĭn and A. A. Migdal, in *Problems of Cybernetics* [in Russian], Vol. 107, R. E. Sagdeev and A. A. Migdal, eds., Scientific Council on Problems of Cybernetics, Moscow (1987), p. 114.

Translated by D. H. McNeill
Edited by P. Shelnitz