Composition of cosmic rays accelerated in supernova remnants

E. G. Berezhko*) and L. T. Ksenofontov†)

Institute of Space Research and Aeronomy, Siberian Branch of the Russian Academy of Sciences, 677891 Yakutsk, Russia

(Submitted 6 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 737–759 (September 1999)

The kinetic theory of regular acceleration of cosmic rays in supernova remnants is used to investigate the expected chemical composition of the rays. It is shown that the shapes of the calculated profiles of the chemical elements making up the cosmic rays are consistent with experiment wherever the results of measurements are available. The acceleration process is accompanied by relative enrichment of the cosmic rays with heavy elements. If the analogous property of the mechanism underlying the injection of superthermal particles into the acceleration regime is taken into account, such enrichment supports the formation of the required composition of cosmic rays in the energy range up to 10^{14}–10^{15} eV. © *1999 American Institute of Physics.* [S1063-7761(99)00109-2]

1. INTRODUCTION

The chemical composition of cosmic rays carries direct information about their sources, the medium from which they originate, and the mechanism of their acceleration. The massive quantity of experimental data accumulated to date suggests that cosmic rays originate in a medium of normal chemical composition typical of the interstellar medium (see, e.g., Ref. 1). This consideration provides further argument to the effect that the bulk of the cosmic rays are produced by shock waves from flares of supernovae, which constitute a natural class of galactic objects capable of providing the energy input required for cosmic rays in the interstellar medium.[1]

On the other hand, there are systematic differences in the compositions of cosmic rays and the interstellar medium. The main difference is that the relative content of nuclei of elements heavier than helium in cosmic rays is higher than in the interstellar medium, and the disparity between them increases as the mass number of the elements A increases. This fact is usually attributed to a necessary property of the cosmic ray acceleration mechanism, that heavier elements are accelerated more efficiently than lighter elements.

Detailed measurements of the composition of cosmic rays are currently being carried out over a wide range of energies, and major efforts are underway to extend this range to the kink region in the cosmic ray spectrum: 10^{15}–10^{16} eV (Refs. 2 and 3). Today it can be said that the demands imposed on cosmic ray sources require not only that they maintain the necessary acceleration efficiency, but also that they generate the observed composition of these rays. Detailed studies of the cosmic ray acceleration process in supernova remnants, based on the nonlinear kinetic theory of acceleration (the theoretical foundations of regular acceleration are presented in two surveys[4,5]), have shown that the acceleration process is characterized by the necessary efficiency: approximately half the energy released in explosion is transferred to accelerated cosmic rays.[6–9]

In this paper we compare calculations of the expected spectra of various cosmic ray components carried out using the nonlinear kinetic theory of cosmic ray acceleration in supernova remnants with existing experimental data, our objective being to determine the extent to which the chemical composition of cosmic rays accelerated in supernova remnants satisfy the experimental requirements.

2. THEORY

The mechanical energy E_{sn} released in a supernova explosion in the initial stage of evolution is represented by the kinetic energy of the expanding shell of ejected matter. The motion of the ejected shell generates a powerful shock wave in the surrounding interstellar medium, the radius of the wave R_s increasing at the rate $V_s = dR_s/dt$. The diffusive propagation of high-energy charged particles in the collisionless dissipative medium enables them to cross the shock front many times before they finally drift into the region behind the shock wave. Every pair of consecutive crossings of the shock front is accompanied by an increase in the particle's energy, wherein lies the essential physical nature of the regular acceleration process.[10–12] In the linear approximation this process generates a power-law cosmic ray momentum spectrum at a planar shock front. Owing to the high acceleration efficiency of acceleration and the hardness of the spectrum, the accelerated cosmic rays constitute an important dynamical factor. The structure of the shock wave modified by the reciprocal influence of cosmic rays on the medium contains—in addition to the usual thermal front—a smooth extended zone known as the prefront. The modification of the shock wave, in turn, is reflected in the shape of the cosmic ray spectrum to the extent that it is not a pure power-law function of momentum in this case. The self-consistent cosmic ray spectrum can be determined by solving the dynamical equations describing their transport and acceleration, along with the dynamics of the medium (gas) with allowance for the reciprocal influence of the rays on the me-

dium; this consideration is the foundation of the kinetic theory of cosmic ray acceleration in supernova remnants as developed in previous studies[6–9] in application to the case in which only protons—the principal species of ions in the interstellar medium–are assumed to be accelerated.

In the kinetic approach the acceleration of cosmic rays by a supernova remnant shock wave is described assuming spherical symmetry by the diffusive transport equation for the cosmic ray distribution function $f_A(r,p,t)$:[13,14]

$$\frac{\partial f_A}{\partial t}=\nabla(\kappa_A\nabla f_A)-\mathbf{w}_c\cdot\nabla f_A+\frac{\nabla\mathbf{w}_c}{3}p\frac{\partial f_A}{\partial p}+Q_A, \qquad (1)$$

where r, t, and p are the radial coordinate, the elapsed time from the supernova remnant explosion, and the particle momentum, respectively, κ_A is the cosmic ray diffusion coefficient, the subscript A designates the nuclear species characterized by the mass number (atomic) A, w_c ($w_c=w$ for $r<R_s$, and $w_c=w+c_a$ for $r>R_s$) is the velocity of the scattering centers, the role of which is taken by Alfvén waves, w is the velocity of the medium, and c_a is the velocity of the Alfvén waves generated by accelerated cosmic rays. In the preshock region, $r>R_s$, the Alfvén waves propagate away from the shock front relative to the medium; in the postshock region they are assumed to be isotropic with respect to the direction of propagation (see, e.g., Ref. 15).

Cosmic rays originate at the thermal front, at which the gas heats up, whereupon a small fraction of the highest-energy particles is capable of crossing the thermal front repeatedly and thus becoming involved in the acceleration process. The injection of particles into the acceleration regime is described by the source

$$Q_A=\frac{N_{\rm inj}^A u_1}{4\pi(p_{\rm inj}^A)^2}\delta(r-R_s)\,\delta(p-p_{\rm inj}^A), \qquad (2)$$

where $u=V_s-w$, $N_{\rm inj}^A$ is the number of gas particles drawn into the acceleration process from unit volume intersecting the thermal front, and $p_{\rm inj}^A$ is the momentum of these particles. The thermal shock front, situated at $r=R_s$, is treated as a discontinuity. The subscript 1 (2) corresponds to the point $r=R_s+0$ ($r=R_s-0$) directly ahead of (behind) the thermal front.

Unfortunately, a systematic theory of collisionless shock transfer in a gas (or, in our terminology, a thermal front) with the capability of reliably predicting the rate of particle injection into the acceleration regime is still lacking. Consequently, the rate of injection of protons as the principal species of ions in the cosmic plasma is specified by the dimensionless parameter[16,17,6–9]

$$\eta=N_{\rm inj}/N_1, \qquad (3)$$

which characterizes the number of injected protons $N_{\rm inj}$, and by the parameter $\lambda>1$, which characterizes their velocity:

$$v_{\rm inj}=\lambda c_{s2}. \qquad (4)$$

Here c_{s2} is the sound velocity in the postshock region, $N_1=\sigma_p N_H$ is the density of protons ahead of the thermal front, N_H is the density of protons in the interstellar medium, and σ_p is the compression ratio of matter within the limits of the prefront. The quantity $v_{\rm inj}$ has the significance of the minimum velocity necessary for particles to be able to cross the shock front and hence enter into the acceleration process. Beginning with the velocity $v_{\rm inj}$, the character of the spectrum of particles is governed by the buildup of energy due to their quasi-cyclic crossing of the shock front, i.e., by regular acceleration.

Note that the total modification of the shock wave to the point where the thermal front completely vanishes[18] is impossible in the presence of an expanding shock wave of finite dimensions,[6–9,19] owing to the influence of geometrical factors.[20]

Inasmuch as the accelerated particles in the shock transition region are powerful energizers of Alfvén waves,[21,22] the diffusion of cosmic rays is assumed to be of the Bohm type, corresponding to the diffusion coefficient

$$\kappa_A=\kappa_B=\rho_B v/3, \qquad (5)$$

where $\rho_B=pc/QeB$ and v are the particle gyroradius and velocity, B is the magnetic field, c is the speed of light, e and m are the proton charge and mass, and Q is the ion charge number. In the perturbation region the diffusion coefficient is $\kappa_A=\kappa_{A0}\rho_0/\rho$, where ρ is the density of the medium, and the subscript 0 refers to the unperturbed interstellar medium. The factor ρ_0/ρ is included, in effect, to account for the postulated increase in the level of turbulence in the perturbed medium and to prevent the development of prefront instability,[23–25] which makes it difficult to solve the stated problem numerically.

In the initial period of acceleration the ion charge number Q is equal to its equilibrium value in the interstellar medium, Q_0. With the passage of time the charge of the ions increases as a result of their ionization due to collisions, which causes the charge number to attain a value equal to the charge number of the nucleus Z. For simplicity we assume that the charge number Q increases in the energy interval $10Amc^2-10^3Amc^2$, directly as the logarithm of the energy, from Q_0 to Z.

The medium (gas) is described by the gasdynamic equations

$$\frac{\partial\rho}{\partial t}+\nabla(\rho\mathbf{w})=0, \qquad (6)$$

$$\rho\frac{\partial\mathbf{w}}{\partial t}+\rho(\mathbf{w}\nabla)\mathbf{w}=-\nabla(P_c+P_g), \qquad (7)$$

$$\frac{\partial P_g}{\partial t}+(\mathbf{w}\nabla)P_g+\gamma_g(\nabla\mathbf{w})P_g=\alpha_a(1-\gamma_g)c_a\nabla P_c, \qquad (8)$$

where ρ, γ_g, and P_g are the density, adiabatic exponent, and pressure of the gas, respectively,

$$P_c=\frac{4\pi c}{3}\sum\int_0^\infty dp\,\frac{p^4 f_A}{\sqrt{p^2+(Amc)^2}} \qquad (9)$$

is the cosmic ray pressure, in which the summation is carried out over all species of nuclei under consideration, $\alpha_a=1$ for $r>R_s$, and $\alpha_a=0$ for $r<R_s$. The reciprocal dynamical influence of cosmic rays on the medium is taken into account

by the term $-\nabla P_c$. According to Eq. (7), the pressure gradient of the cosmic rays directly influences the velocity profile of the medium $w(r)$. The influence of the cosmic rays is particularly strong in the preshock region, $r>R_s$, where their pressure varies considerably within a relatively small space scale $(10^{-3}-10^{-2})R_s$ (Ref. 8), resulting in significant modification of the shock transition, i.e., the formation of a prefront.

The influence of the cosmic ray pressure is also substantial in the postshock region, $r<R_s$, even though it does not produce the same radical changes therein, because its gradient is much weaker.

In addition, the cosmic ray pressure enters into Eq. (8) and induces considerable additional heating of the gas in the region $r>R_s$. The role of cosmic rays is intermediate in this case: They strongly energize Alfvén waves in the region $r>R_s$, whose subsequent dissipation has the effect of heating the gas.[22] In the case of a high acoustic Mach number $M_s=V_s/c_{s0}\gg\sqrt{M_a}$, where $M_a=V_s/c_{s0}$ is the Alfvén Mach number, and the subscript 0 refers to the unperturbed medium, the dissipation of Alfvén waves has a strong influence on the structure of the modified shock wave. It severely limits the growth of the compression ratio $\sigma=\rho_2/\rho_0$ of the matter at the level $\sigma\approx M_a^{3/8}$, whereas ignoring this effect implies $\sigma\approx M_s^{3/4}$ (Refs. 7 and 8). Since the Alfvén Mach number is high in the active period of evolution of the shock wave, $M_a\sim M_s\gg 1$, when the bulk of the cosmic rays are produced, the compression ratio σ far exceeds the classical limit 4 in this case.

For a stationary, homogeneous medium, which is the only case treated below, Eqs. (1) and (6)–(8) are solved subject to the initial $(t=0)$ and boundary $(r=\infty)$ conditions

$$f_A(p)=0, \quad \rho=\rho_0, \quad P_g=P_{g0}, \quad w=0. \tag{10}$$

We note that the condition $f_A(r=\infty,p,t)=0$ implies neglect of the role of background cosmic rays, which is justified because injection and subsequent entrainment of superthermal gas particles in the acceleration regime dominates preaccelerated background cosmic rays, owing to the low density of the latter.

In the very earliest period of expansion of the explosion-ejected shell its matter is described by a self-similar distribution in terms of the velocities $v=r/t$ (see Refs. 26–28):

$$\rho_{\rm ej}=\begin{cases} Ft^{-3}, & v<v_t, \\ Ft^{-3}(v/v_t)^{-k}, & v\geqslant v_t, \end{cases} \tag{11}$$

where

$$F=\frac{1}{4\pi k}\frac{[3(k-3)M_{\rm ej}]^{5/2}}{[10(k-5)E_{\rm sn}]^{3/2}}, \quad v_t=\left[\frac{10(k-5)E_{\rm sn}}{3(k-3)M_{\rm ej}}\right]^{1/2},$$

$M_{\rm ej}$ is the mass of the shell, and the parameter k has a value typically in the interval between 7 and 12.

The internal pressure in the shell matter is negligible. Interaction with the environment modifies the density distribution and produces a backward shock wave inside the ejected matter; this wave, in turn, heats the gas to very high temperatures (see, e.g., Ref. 29). In this paper we take the backward shock wave into account and use an approximate description of the ejected matter dynamics, representing it by two constituent parts: a thin shell (the piston) moving with a certain velocity V_p, and a free expanding part, which is described by the distribution (11). The piston consists of the hindered part of the distribution (11) with initial velocities $v>R_p/t$:

$$M_p=4\pi t^3\int_{R_{p/t}}^{\infty} dv\, v^2\rho_{\rm ej}(v,t), \tag{12}$$

where R_p is the radius of the piston. The dynamics of the piston is satisfactorily described by the thin-shell approximation,[30] whereby the motion of the piston obeys the equation[9]

$$\frac{d(M_pV_p)}{dt}=4\pi R_p^2[P_c(r=R_p-0) - P_g(r=R_p+0)-P_c(R_p+0)]. \tag{13}$$

The cosmic ray pressure $P_c(r=R_p-0)$ is created by particles that have penetrated the piston. The boundary condition on the surface of the piston characterizes the balance of diffusion fluxes of cosmic rays across the boundary $r=R_p$:

$$\kappa\frac{\partial f_A}{\partial r}=-\phi_A \quad \text{at} \quad r=R_p+0, \tag{14}$$

where

$$\phi_A=\kappa_A[f_A(r=R_p-0,p,t)-f_A(r=R_p+0,p,t)]/l_p,$$

$l_p=\delta R_p$ is the thickness of the piston, and the value $\delta=0.1$ is used in the calculations.

In the region behind the piston $(r<R_p)$ the cosmic ray distribution function is described by the approximate equation

$$\frac{\partial f_A}{\partial t}=\frac{V_p}{R_p}p\frac{\partial f_A}{\partial p}-\frac{S}{V}\phi_A, \tag{15}$$

which follows from Eq. (1) with allowance for the fact that, owing to the large diffusion coefficient, the spatial distribution of the cosmic rays in the volume $V=4\pi R_p^3/3$ is almost homogeneous.

The penetration of cosmic rays through the piston is not a significant factor from the standpoint of the evolution of the shock wave or the acceleration of the rays themselves.[7,8] In the early stage of free expansion this process is inconsequential and the diffusion coefficient κ_A is therefore small, owing to the high density ρ_p. In the Sedov intermediate stage, when the majority of the cosmic rays are generated, the radius of the piston is small in comparison with that of the shock wave $(R_p\ll R_s)$. Consequently, in this stage of evolution of the supernova remnant as well, the penetration of cosmic rays into the region $r<R_p$ does not exert a major influence on the shock wave dynamics or the acceleration of the rays.

The investigation of the real distribution (11) of ejected matter is important in the initial stage of evolution of the supernova remnant, when the mass of matter swept from the surrounding medium is much smaller than the mass of ejected matter $M_{\rm ej}$ (Ref. 9). During this period the velocity of

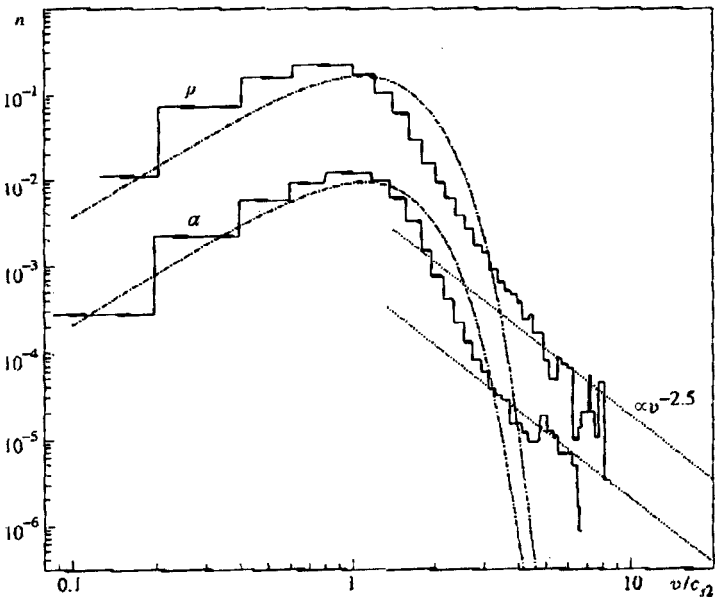

FIG. 1. Differential densities of protons and alpha particles behind the front of a longitudinal shock wave with Mach number $M_s = 7.1$ and Alfvén Mach number $M_a = 4.6$ as a function of particle velocity. The histogram is the result of numerical simulation;[33] the dot–dash curves correspond to a Maxwellian distribution with parameters corresponding to the Rankine–Hugoniot equations; the points represent power-law spectra corresponding to the theory of regular acceleration.

the piston V_p is much higher than the average $V_{p0} = \sqrt{2E_{\rm sn}/M_{\rm ej}}$, so that the shock velocity is higher, and the acceleration of the cosmic rays is more rapid than in the frequently used simplified case where, in contrast with (11), the entire shell matter is assumed to expand with the velocity V_{p0}.

We disregard the influence of radiation cooling of the perturbed medium. This process becomes important in the later stage of evolution of the supernova remnant,[31] when the acceleration of cosmic rays is rendered inefficient by the low velocity of the shock wave.

This problem is exactly solvable only by numerical methods. A description of the fundamentals of a newly developed and implemented numerical algorithm can be found in previous papers,[6–9] along with a detailed discussion of the physical results obtained via the algorithm for the case in which only protons—as the principal species of ions in the interstellar medium—are taken into account.

The results of a numerical simulation of collisionless, quasi-longitudinal shock waves,[32,33] experimental results obtained in interplanetary space,[34,35] and the results of a recently developed analytical theory[36,37] show that superthermal particles can be efficiently injected into the acceleration regime at a rate exceeding $\eta = 10^{-4}$.

Calculations have shown[8] that the indicated injection rate makes the acceleration very efficient: Even at $\eta = 10^{-4}$ more than 20% of the explosion energy $E_{\rm sn}$ is transferred to cosmic ray protons. An increase in the injection rate causes the cosmic ray acceleration efficiency to increase. However, this functional dependence is not very strong; increasing η by two orders of magnitude produces at most a threefold increase in the cosmic ray energy content. It is safe to say that the actual value of η is not too critical from the energy standpoint.

The situation changes dramatically when it becomes necessary to investigate the expected chemical composition of accelerated cosmic rays. In the interstellar medium the nuclei of all elements constitute small impurities in relation to protons.[1] Allowance for their acceleration cannot significantly influence the shock structure and dynamics. Consequently, the spectra of these elements can be calculated in the linear approximation. This means that the number of accelerated particles of an element with mass number A is directly proportional to the number of particles involved in acceleration, $N^A_{\rm inj}$, which, in turn, is expressed in terms of the injection parameter

$$\eta_A = N^A_{\rm inj}/N_{A1}, \tag{16}$$

where $N_{A1} = \sigma_p N_A$, and N_A is the density of the element with mass number A in the interstellar medium. Consequently, the number of accelerated nuclei with mass number A is proportional to the parameter η_A. In other words, unlike the acceleration efficiency, the resultant composition of cosmic rays is in fact sensitive to the injection rate or, more precisely, to its dependence on the particle species.

Numerical simulation[33] and experimental[38] results show that the thermalization of the gas at the shock front produces an approximately identical velocity distribution for all ion species. This fact is demonstrated in Fig. 1, which shows the differential spectra (derivative of the number density with respect to the velocity v) $n = dN/dv$ of protons and α-particles at the shock front, plotted from the results of numerical simulation of a longitudinal shock wave with Mach number $M_s = 7.1$ (Ref. 33). This result is consistent with the notion that the conversion of energy of directional motion of the gas impinging on the shock front is a consequence of elastic scattering of gas particles by inhomogeneities of the magnetic field frozen into the medium.

If the injection mechanism selected particles by their velocities, the momenta of injected protons $p_{\rm inj}$ and heavier elements $p^A_{\rm inj}$ would be related by the equation

$$p^A_{\rm inj} = A p_{\rm inj}, \tag{17}$$

where the fraction of particles of different species involved in acceleration would be the same, $\eta_A = \eta$.

However, calculations and experiment both indicate that the acceleration process involves a much smaller fraction of particles of the thermal distribution than are purely capable kinematically of moving out of the postshock region, crossing the shock front, and thus entering into the acceleration process. This means that within the shock transition zone there are more significant factors affecting the particles and enabling particles of insufficient velocity to cross the shock front. The basis of these factors is the interaction of particles with the electromagnetic field. Particles are therefore selected according to their rigidities: The acceleration process involves particles whose hardness $R \propto p/Q$ (Q is the ion charge number) exceeds a certain critical value $R_{\rm inj}$. Consequently, the velocity of injected heavier elements $v_{\rm inj}^A = v_{\rm inj}/Q$ is lower than the proton velocity $v_{\rm inj}$. Since the power-law spectrum of the accelerated particles matches up with the much steeper thermal spectrum at the point $v_{\rm inj}^A$, the fraction of injected particles can be expected to be an increasing function of the dimensionless parameter A/Q; this function can be formally represented as

$$\eta_A = \eta e_{\rm inj}(A/Q), \tag{18}$$

where the factor $e_{\rm inj}(x) \geqslant 1$ is a monotonically increasing function of its argument $x \geqslant 1$. The same is true if the number of injected particles is converted to the injection momentum $p_{\rm inj}^A = A p_{\rm inj}$, which we have used in the calculations as a matter of convenience. This assertion is evident, in particular, from Fig. 1, which shows (with a certain indeterminacy due to the increased statistical error of the calculations as the velocity increases) that the actual injected proton velocity is $v_{\rm inj} \simeq 6 c_{s2}$, because, beginning with the velocity $v = 6 c_{s2}$, the proton spectrum obeys a power law, $n \propto v^{-\gamma}$, whose exponent is close to the universal value $\gamma = (\sigma_{\rm ef}+2)/(\sigma_{\rm ef}-1) = 2.5$, where $\sigma_{\rm ef} = \sigma(1-1/M_a) = 2.95$ is the effective compression ratio, $M_a = 4.6$ is the Alfvén Mach number, and $\sigma = 3.78$ is the compression ratio corresponding to the acoustic Mach number $M_s = 7.1$. On the other hand, in accordance with the role of hardness, the velocity of injected α-particles is approximately half the proton velocity, $v_{\rm inj}^{\alpha} \simeq 3 c_{s2}$. But then, if $v_{\rm inj} = 6 c_{s2}$ is assumed for particles of both species, the α-particle enrichment factor is $e_{\rm inj}(2) \simeq 2$.

It should also be noted, as is evident from Fig. 1, that the profile of the self-consistent spectrum of thermal particles differs significantly from the Maxwellian form $n \propto v^2 \times \exp(-v^2/v_T^2)$ $v_{\rm inj}$, even at velocities well below the threshold value $v > v_{\rm inj}$.

Since, as was mentioned, the dynamical role of elements heavier than hydrogen is of minor significance, it follows that the efficiency of cosmic ray acceleration, the salient features of the modification of the shock wave, and the evolution of the latter do not differ markedly from the case when only the acceleration of protons is taken into account. In the present discussion, therefore, we shall not consider these aspects of the problem in detail, referring the reader instead to previous papers.[6–9] The ensuing discussion will be concerned primarily with the properties of the resultant energy spectrum of cosmic rays produced by the shock wave during the entire period of its active (from the standpoint of cosmic ray production) evolution:

$$N_c^A(\epsilon_k) = \frac{16\pi^2(\epsilon_k + Amc^2)p}{c} \int_0^{\infty} dr r^2 f_A(r,p,t). \tag{19}$$

Here $\epsilon_k = \epsilon - Amc^2$ is the kinetic energy of a particle with mass number A and ϵ is its total energy, which is related to the momentum by the usual equation $\epsilon = \sqrt{p^2c^2 + (Amc^2)^2}$. The spectrum $N_c^A(\epsilon_k)$ varies in the course of evolution, particularly in its early stages, including the free expansion stage and the initial Sedov stage, when the amplitude of the spectrum increases as an ever-greater number of particles enters into the acceleration process, and the maximum particle energy $\epsilon_{\rm max}$, at which an abrupt exponential cutoff of the spectrum $N_c^A(\epsilon_k)$ takes place, increases as well. In the later stages of evolution, when the Mach number of the shock wave decreases to $M_s \approx 4$ as a result of its slowing down, the acceleration of cosmic rays becomes relatively inefficient, and the spectrum $N_c^A(\epsilon_k)$ remains essentially unchanged. Consequently, the term ''resultant'' refers to the spectrum calculated from Eq. (19) for the indicated later stage of evolution.

3. RESULTS OF CALCULATIONS AND DISCUSSION

A large portion of the volume of the Galaxy is occupied by the so-called hot phase and warm phase of the interstellar medium[1] with densities of hydrogen atoms, temperatures, and magnetic fields $N_H = 0.003\,{\rm cm}^{-3}$, $T_0 = 10^6\,{\rm K}$, $B_0 = 3\,\mu{\rm G}$ and $N_H = 0.3\,{\rm cm}^{-3}$, $T_0 = 10^4\,{\rm K}$ $B_0 = 5\,\mu{\rm G}$, respectively. We therefore confine our discussion to the evolution of supernova remnants in the two types of interstellar medium.

The evolution of supernova remnants in a homogeneous interstellar medium is characteristic of type Ia supernovae, for which the modification of the surrounding interstellar medium by the stellar wind is insignificant.[39] In the calculations we use typical parameters of type Ia supernovae: explosion energy $E_{\rm sn} = 10^{51}\,{\rm erg}$, mass of the ejected shell $M_{\rm ej} = 1.4 M_{\odot}$ and $k = 7$.

According to calculations,[40] the charge number of ions in the hot interstellar medium increases smoothly from $Q_0 = 1$ for hydrogen (H) to $Q_0 = 8$ for iron (Fe). In a warm interstellar medium the ion charge is assumed to be $Q_0 = 1$ for all elements.

The density $N_A = N_{\rm H} a_A$ of elements heavier than hydrogen in the interstellar medium is assumed to be proportional to the relative (to hydrogen) abundance a_A of the element in the local region of the Galaxy.

The numerical solution of the problem of the evolution of the shock wave generated by the explosion-ejected shell sets the stage for calculating the resultant cosmic ray spectrum created by it, $N_c^A(\epsilon_k)$, i.e., the differential (with respect to the kinetic energy ϵ_k) number of accelerated rays. Proceeding from the knowledge that supernova remnants are the primary source of cosmic rays (at least up to energies $\epsilon_k \sim 10^{14}$–$10^{15}\,{\rm eV}$), we can determine the differential intensity of cosmic rays in the Galaxy with allowance for their escape (or the finiteness of their residence time in the Galaxy):

$$J_A^G(\epsilon_k) \propto \tau_e(R) v N_c^A(\epsilon_k). \tag{20}$$

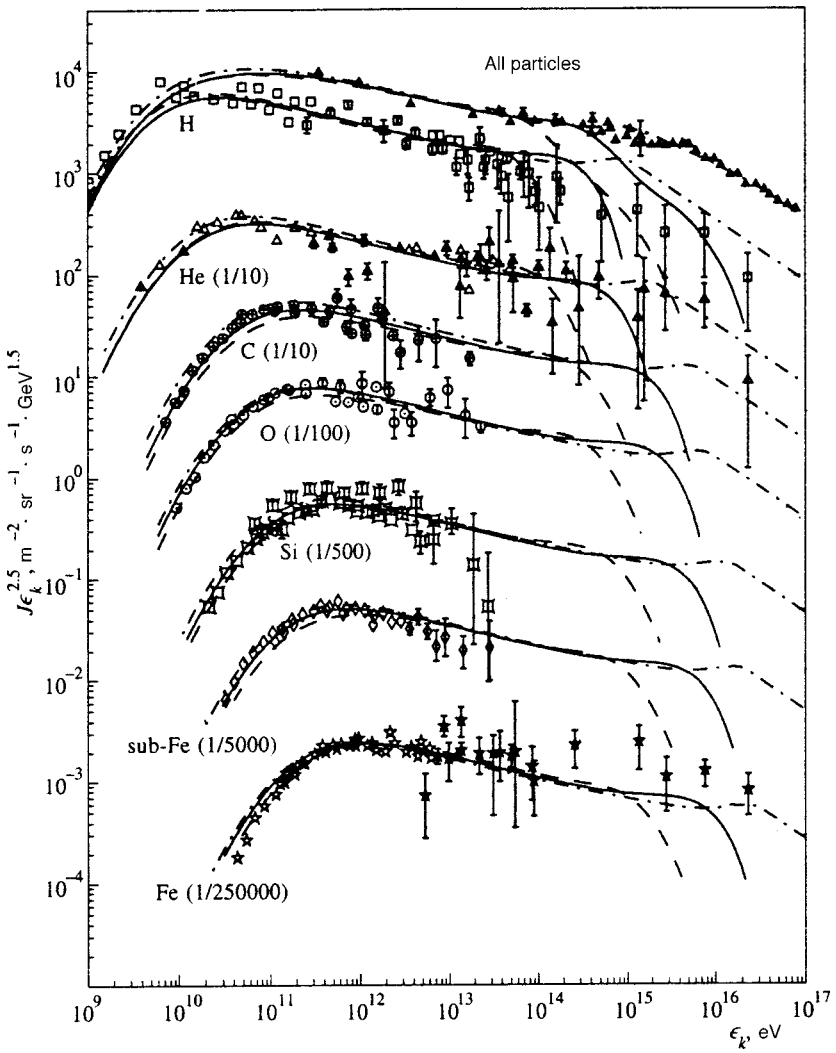

FIG. 2. Intensity of cosmic rays on Earth as a function of the kinetic energy. The experimental values are taken from a survey paper.[2] The solid (dashed) curves represent calculations for the hot (warm) phase of the interstellar medium at an injection rate $\eta=10^{-4}$. The dot–dash curves correspond to a hot interstellar medium with a magnetic field $B_0=12\,\mu$G at an injection rate $\eta=5\times10^{-4}$ with the spectra extended into the high-energy range.

Here τ_e is the average residence time of cosmic ray particles in the Galaxy, which is a function of their hardness $R(\epsilon_k)$. The profile of the cosmic ray spectrum at the source, $N_c^A(\epsilon_k)$, and the profile of their average spectrum in the Galaxy, $J_A^G(\epsilon_k)$, differ, because the residence time of cosmic rays in the Galaxy τ_e depends significantly on their energy (or, more precisely, their hardness).

The functional dependence $\tau_e(R)$ determined on the basis of the measured ratio of the primary (generated at the source) and secondary (resulting from the interaction of primary cosmic rays with nuclei of the interstellar medium) components of the cosmic rays can be written in the form

$$\tau_e \propto (R_0+R)^{-\mu}, \tag{21}$$

where $R_0=5$ GV and $\mu=0.3$–0.7 (Ref. 1).

The cosmic ray spectrum observed on Earth $J_A(\epsilon_k)$ differs from the average spectrum in the Galaxy $J_A^G(\epsilon_k)$ because of the modulating effect of the solar wind, which lowers the cosmic ray intensity $J_A(\epsilon_k)$ in the low-energy region relative to $J_A^G(\epsilon_k)$. This effect can be described quantitatively on the assumption that each particle reaching Earth's orbit loses, on the average, an amount of energy

$$\Delta\epsilon = Ze\phi,$$

which is determined by the value of the potential ϕ (Ref. 41). In this case the cosmic ray intensity observed on Earth is expressed in terms of their intensity outside the heliosphere by the relation

$$J_A(\epsilon_k)=\left(\frac{\epsilon^2-\epsilon_0^2}{\epsilon^2+2\epsilon\Delta\epsilon+\Delta\epsilon^2-\epsilon_0^2}\right)J_A^G(\epsilon_k+\Delta\epsilon), \tag{22}$$

in which $\epsilon=\epsilon_k+\epsilon_0$ is the total energy of the particle, and $\epsilon_0=Amc^2$ is its rest energy.

In Fig. 2 the expected terrestrial cosmic ray spectra from supernova remnants, $J_A(\epsilon_k)$, calculated for a number of elements, are compared with existing experimental data. The calculations correspond to a moderate injection rate $\eta=10^{-4}\,\tau_e$. It is important to note that the uncertainty of the supernova parameters and the residence time of cosmic rays in the Galaxy prevents the expected cosmic ray intensity from being calculated with the required accuracy, even for protons. Consequently, the normalization of the theoretical spectrum of the proton component of the cosmic rays is chosen in such a way as to achieve the best agreement with experiment.

The injection rates η_A or, equivalently, the factors $e_{\rm inj}(A/Q_0)$, are chosen in such a way that, given the same

normalization as for protons, the amplitude of the spectra $J_A(\epsilon_k)$ of all other elements will agree with experiment at the energy $\epsilon_k \sim 10^{10} A$ eV. The spectrum of all particles

$$J_\Sigma(\epsilon_k) = \sum J_A(\epsilon_k)$$

is calculated by summing the element spectra shown in Fig. 2.

In the region $\epsilon_k \lesssim A$ GeV, where the modulation of cosmic rays by the solar wind becomes important, the calculated spectra are made to agree with experiment by choosing the parameter $\phi = 600$ MV, which governs the energy losses by cosmic ray particles $\Delta\epsilon = Ze\phi$ when they penetrate Earth's orbit from the interstellar medium.

The main conclusion that can be drawn from Fig. 2 is that the existing experimental data on the contents of various elements in the composition of cosmic rays are consistent, in terms of the profile of the spectrum, with the predictions of the theory of their acceleration in supernova remnants in the energy range $\epsilon_k \lesssim 10^{15}$ eV. It is important to note that calculations corresponding to different phases of the interstellar medium exhibit equally good agreement with experiment. Consequently, the implementation of the necessary procedure for averaging the resultant cosmic ray spectrum with allowance for the allocation of supernova remnants between the phases of the interstellar medium (for which data are all but nonexistent) cannot influence the conclusion as to the degree of agreement between theory and experiment.

The spectrum of protons as the main component of cosmic rays in the given energy range warrants special attention. The experimental cosmic ray fluxes shown in Fig. 2 have been obtained in different experiments.[2] An analysis of the entire set of existing measurement data reveals good agreement with a power-law proton spectrum $J(\epsilon_k) \propto \epsilon_k^{-\gamma}$ with a single exponent $\gamma = 2.75$ over the entire range $\epsilon_k \lesssim 10^{15}$ eV. On the other hand, an analysis of data from individual experiments leads to a significantly different conclusion, that the proton spectrum has a kink and becomes considerably softer at $\epsilon_k \sim 1$ TeV (Ref. 42). There is no simple explanation for this situation from the theoretical point of view, because, for example, prominent features are not observed in the helium spectrum at these energies. Additional experiments planned for the very near future will doubtless shed light on this dilemma.

The energy limit of cosmic rays accelerated in supernova remnants is dictated by geometrical factors and for the postulated Bohm diffusion coefficient of cosmic rays (5) can be estimated from the expression[20]

$$\epsilon_{\max} = 5\times 10^{14} Z \left(\frac{E_{\rm sn}}{10^{51}\,{\rm erg}}\right)^{1/2} \left(\frac{M_{\rm ej}}{1.4 M_\odot}\right)^{-1/6} \times \left(\frac{N_H}{3\times 10^{-3}\,{\rm cm}^{-3}}\right)^{1/3} \left(\frac{B_0}{3\,\mu{\rm G}}\right) {\rm eV}, \quad (23)$$

which is in good agreement with the results of calculations in Fig. 2. We note that Eq. (23) is directly applicable to the spectra of separate elements $J_A(\epsilon_k)$. As for the spectrum of all particles $J_\Sigma(\epsilon_k)$, it is evident from Fig. 2 that it is characterized by the maximum energy $\epsilon_{\max}$, which is only slightly higher than the maximum proton energy.

The following consideration is of utmost importance. At a low injection rate the spectrum of cosmic rays accelerated in supernova remnants is close to $N_c \propto \epsilon_k^{-2}$ in the relativistic range. Increasing the injection rate enhances the modification of the shock wave and the hardness of the resultant cosmic ray spectrum. The case represented in Fig. 2 corresponds to moderate injection. Agreement with experiment is achieved for $\mu = 0.73$ in the hot phase and for $\mu = 0.86$ in the warm phase of the interstellar medium, both of these values lying outside the experimentally established interval $\mu = 0.3$–0.7 (Ref. 1). In other words, the theory predicts a somewhat harder cosmic ray spectrum than is required. This contradiction could be resolved by the preacceleration of cosmic rays. Since the preacceleration mechanism is less efficient than for regular acceleration (as we must assume to be true, otherwise regular acceleration could not be regarded as the primary process generating the spectrum of cosmic rays), it will affect low-energy particles to a greater degree. This fact implies a certain increase in the amplitude of the cosmic ray spectrum, the amount of the increase being a decreasing function of the cosmic ray energy, so that softening of the spectrum is implied. The preacceleration of cosmic rays can be achieved both by statistical mechanisms and as a result of their repeated interaction with the whole ensemble of existing shock waves in the Galaxy. Inasmuch as the probability of repeatedly encountering a shock front is proportional to the number of shock waves, cosmic ray particles most often interact with weak shocks, which can make their resultant spectrum softer.

It must be borne in mind, however, that this is not a simple problem, even if one proceeds from energy considerations alone. The transformation of the cosmic ray spectrum $N_c(\epsilon) \propto \epsilon^{-2}$ in the range 10^9 eV$\leqslant \epsilon \leqslant 10^{14}$ eV into the spectrum $N_c(\epsilon) \propto \epsilon^{-2.1}$ without any change in the amplitude at the maximum energy $\epsilon_{\max} = 10^5$ GeV requires the expenditure of higher energies than that contained in the initial cosmic ray spectrum.

To illustrate the enrichment of cosmic rays with heavy elements, which is maintained by regular acceleration, in Fig. 3 the calculated enrichment factor

$$e(\epsilon_k/A) = [I_A(\epsilon_k/A)/I_H(\epsilon_k/A)/(N_A/N_H)] \quad (24)$$

at the energy $\epsilon_k/A = 3$ GeV/A is compared with the experimental data,[43] where

$$I_A(\epsilon_k/A) = A J_A(\epsilon_k) \quad (25)$$

is the differential (with respect to the kinetic energy per nucleon) cosmic ray intensity. The calculations have been carried out for three different values of the injection rate η: 10^{-5} (low rate); 3×10^{-4} (moderate rate); 10^{-3} (high rate).

It is evident from Fig. 3 that for a low injection rate, when cosmic rays are accelerated by the unmodified shock wave, enrichment does not take place during acceleration: $e < 1$ holds for all elements with $A > 1$. An increase in the

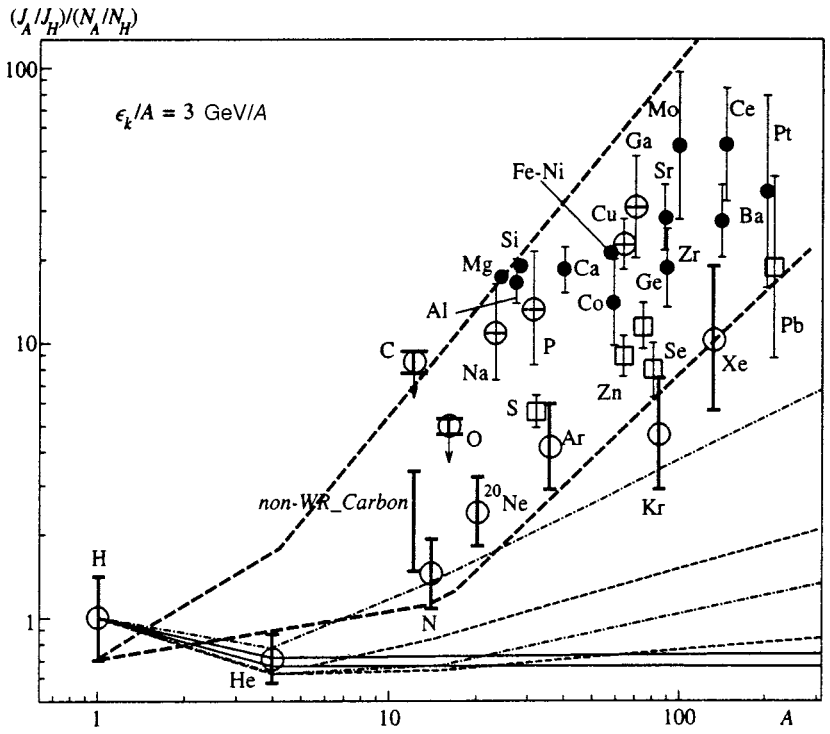

FIG. 3. Cosmic ray enrichment factor at energy $\epsilon_k/A = 3$ GeV/A as a function of the mass number. The experimental values are taken from Ref. 43. The elements denoted by circles, squares, circles with crosses, and filled circles correspond to condensation temperatures $T_c < 400$ K, $T_c = 400-875$ K, $T_c = 875-1250$ K, and $T_c > 1250$ K, respectively. The solid curves correspond to an injection rate $\eta = 10^{-5}$, the light dashed curves to $\eta = 3\times10^{-4}$, the dot–dash curves to $\eta = 10^{-3}$, and the bold dashed curves to $\eta = 3\times10^{-4}$ with an enrichment factor $e_{\rm inj} = A/Q_0$. For every pair of like curves, the lower one corresponds to the hot interstellar medium, and the upper to the warm phase of the interstellar medium.

injection rate has the effect of increasing the enrichment factor, which is always greater in the warm interstellar medium than in the hot phase.

At a higher injection rate, when modification of the shock wave by the reciprocal influence of cosmic rays becomes significant, the enrichment of cosmic rays with heavy elements becomes more and more important. In other words, this property of regular acceleration has a purely nonlinear character. Its basis is the fact that heavier nuclei have a greater diffusion length $l = \kappa/u \propto (\epsilon_k/A)(A/Q)$ at equal energy per nucleon ϵ_k/A, because we always have $Q < A$. Here u is the velocity of the medium relative to the shock front. In the presence of a modified shock the particles ''sense'' the velocity drop within the limits of the shock transition, $\Delta u = u(l) - u_2$ (u_2 is the velocity of the medium immediately behind the shock front), which is an increasing function of l. For the same energy per nucleon, therefore, heavy elements have a harder spectrum, thus accounting for their higher relative content in the range of relativistic energies.

To estimate numerically the degree of enrichment of cosmic rays with heavy elements produced in the acceleration process, we write the spectrum of accelerated protons in the simplified form

$$N_c \propto \begin{cases} \dfrac{N_{\rm inj}}{p_{\rm inj}}\left(\dfrac{p}{p_{\rm inj}}\right)^{-\gamma_n}, & p \leqslant mc, \\ \dfrac{N_{\rm inj}}{p_{\rm inj}}\left(\dfrac{mc}{p_{\rm inj}}\right)^{-\gamma_n}\left(\dfrac{p}{mc}\right)^{-\gamma_r}, & p \geqslant mc, \end{cases} \tag{26}$$

which explicitly reveals that nonrelativistic protons are accelerated in part of the shock transition—the thermal front—and therefore have a softer spectrum ($\gamma_n > 2$) than relativistic particles, for which $\gamma_r < 2$. The values of the parameters $p_{\rm inj}$, $N_{\rm inj}$, γ_n, and γ_r in Eq. (26) must refer to the most active (in terms of the efficiency with which cosmic rays are generated) phase of evolution of supernova remnants, i.e., the beginning of the adiabatic stage. The exponents of the spectrum are therefore related to the shock parameters in this phase—the compression ratio of matter at the thermal front σ_s and the total compression ratio σ—by the equations[20,44]

$$\gamma_n = (\sigma_s + 2)/(\sigma_s - 1), \tag{27}$$

$$\gamma_r = 1.5 + (3.5 - 0.5\sigma_s)/(2\sigma - \sigma_s - 1). \tag{28}$$

Since $\sigma_s \leqslant 4$ and $\sigma \geqslant 4$ always hold, we have a softer spectrum in the nonrelativistic range ($\gamma_n > 2$) than in the relativistic range ($\gamma_r \leqslant 2$).

The spectrum of heavier elements $N_c^A(p_A)$ is formed under the conditions of the shock transition, whose structure is determined by the modification of protons. It has a kink at the momentum $p_A = Qmc$, which corresponds to the same hardness as exhibited by protons with momentum $p = mc$:

$$N_c^A \propto \begin{cases} \dfrac{N_{\rm inj}^A}{p_{\rm inj}^A}\left(\dfrac{p_A}{p_{\rm inj}^A}\right)^{-\gamma_n}, & p \leqslant Qmc, \\ \dfrac{N_{\rm inj}^A}{p_{\rm inj}^A}\left(\dfrac{Qmc}{p_{\rm inj}^A}\right)^{-\gamma_n}\left(\dfrac{p_A}{Qmc}\right)^{-\gamma_r}, & p \geqslant Qmc. \end{cases} \tag{29}$$

Invoking relations (26)–(29), we can readily estimate the enrichment factor in the relativistic energy range:

$$e = \left(\frac{I_A}{I}\right) \Big/ \left(\frac{N_A}{N_H}\right) = e_{\rm inj}\left(\frac{A}{Z}\right)^{-\mu}\left(\frac{A}{Q}\right)^{\gamma_n - \gamma_r}. \tag{30}$$

It is evident from this expression that the enrichment factor is the product of three factors:

$$e = e_{\rm inj} e_{\rm esc} e_{\rm acc}.$$

The first factor $e_{\rm inj}$ is associated with the injection characteristics. The second

$$e_{\rm esc}=(A/Z)^{-\mu} \tag{31}$$

is attributable to the dependence of the residence time of cosmic rays in the Galaxy on the particle species. It is always less than unity, resulting in the depletion of heavy elements from the cosmic rays. Since the approximate relation $Z \simeq A/2$ holds for all nuclei heavier than protons, the factor

$$e_{\rm esc}\simeq 2^{-\mu} \tag{32}$$

is the same for all elements.

The enrichment of cosmic rays with heavy elements during acceleration is characterized by the factor

$$e_{\rm acc}=(A/Q)^{\gamma_n-\gamma_r}, \tag{33}$$

which depends significantly on the degree of shock modification by the reciprocal influence of accelerated cosmic rays. At a low injection rate the shock is only slightly modified ($\sigma_s\simeq\sigma\simeq 4$), so that the cosmic ray spectrum has a single power-law profile with exponent $\gamma_n\simeq\gamma_r\simeq 2$, and it follows from Eq. (33) that this profile precludes enrichment ($e_{\rm acc}=1$ $e<1$).

For moderate injection ($\eta=3\times10^{-4}$) the degree of modification becomes significant: During the period of most active generation of cosmic rays the shock is characterized by the parameters $\sigma_s=3.4$ and $\sigma=6.7$ for the warm interstellar medium, with $\sigma_s=3.6$ and $\sigma=4.6$ for the hot medium. As a result, the difference in the exponents of the spectra in the relativistic and nonrelativistic ranges is $\gamma_n-\gamma\simeq 0.5$ for the warm interstellar medium and $\gamma_n-\gamma\simeq 0.2$ for the hot medium.

At a high injection rate ($\eta=10^{-3}$) the shock modification is even higher: in the active period $\sigma_s=3.1$, $\sigma=7.1$, and $\gamma_n-\gamma_r=0.7$ for the warm interstellar medium, with $\sigma_s=3.3$, $\sigma=4.8$, and $\gamma_n-\gamma_r=0.4$ for the hot medium.

It is evident from Fig. 3, therefore, that for moderate and high injection rates the factor $e_{\rm acc}$ increases with the atomic number A, its value rising considerably for all $A>1$ as the injection rate η increases. On the other hand, even at a high injection rate and for a warm interstellar medium, where the factor $e_{\rm acc}$ is a maximum, the acceleration process does not support the required degree of enrichment of cosmic rays with heavy elements. As mentioned above, there is every reason to identify this discrepancy between theory and experiment with the heavy-element enrichment of cosmic rays during particle injection into the acceleration regime, i.e., with the factor $e_{\rm inj}(A/Q)>1$.

If we use the results in Fig. 1 as guidelines, we can conclude that the injection process is characterized by the enrichment factor $e_{\rm inj}=A/Q_0$.

Figure 3 shows curves corresponding to an injection rate $\eta=3\times10^{-4}$ with the factor $e_{\rm inj}=A/Q_0$ taken into account. Clearly, the experimental values fall into the region between the two curves for different types of interstellar medium. This fact implies that the regular acceleration of cosmic rays in supernova remnants supports the observed dependence of cosmic ray generation on A, owing to the dependence of the acceleration efficiency and the particle injection rate into the acceleration regime on the hardness of the particles.

We note that the necessary value of the enrichment factor e can be attained exclusively by virtue of the acceleration properties as long as an injection rate higher than $\eta=10^{-3}$ is considered. As was mentioned, however, even for $\eta=10^{-4}$ the calculated cosmic ray spectrum is somewhat harder than required. With a substantial increase in the injection rate the general increase in the hardness of the spectrum in the relativistic energy range of the cosmic ray spectrum is accompanied by the emergence of a bump at energies immediately adjacent to the limit $\epsilon_{\rm max}$ (Ref. 45), in contradiction with experiment. It is most probable, therefore, that the injection rate lies in the range $\eta=10^{-4}$–10^{-3}, where the properties of the injection and acceleration of cosmic rays are capable of reproducing the observed functional dependence $e(A)$.

It is important to note that the observed composition of the cosmic rays, as is evident from Fig. 3, is not reducible to a smooth dependence $e\propto A^{\alpha}$. Significant anomalies are associated with carbon and oxygen nuclei, whose content in cosmic rays is substantially higher than that of adjacent elements. As mentioned in Ref. 46, this characteristic can be attributed to the contribution of type Ib supernovae. Wolf–Rayet stars are presupernovae in this setting; they are characterized by a strong wind, which delivers into the surrounding medium $\sim 10M_{\odot}$ of matter, in whose composition carbon and oxygen are the predominant elements. Consequently, cosmic rays generated by shock waves from the explosion of Wolf–Rayet stars are highly enriched with these elements in comparison with their content in the normal interstellar medium.

The monotonic behavior of the dependence $e(A)$ is also violated by the content of refractory elements in the cosmic ray composition (see Fig. 3). The main portion of these elements exists in the interstellar medium as dust. Consequently, the mechanism of injection into the regular acceleration regime can differ substantially from the usual case of elements that move through the interstellar medium as solitary atoms. It is highly probable that the injection of these elements takes place in two stages.[46] In the first stage, acceleration involves actual grains considerably harder than solitary ions. Upon reaching a certain critical energy, collisions break up the grains into solitary ions, which are drawn directly into the acceleration process. An analysis shows[46] that this phenomenon can account for the observed content of high-melting elements in the cosmic ray composition.

It is clear from the foregoing discussion that supernova remnants (source I) are capable of forming the spectrum of cosmic rays in the range up to $\epsilon_{\rm max}\sim 10^{15}$ eV. It is legitimate, therefore, to inquire how, in its general features at least, the spectrum of cosmic rays is formed at high energies.

Although little is known about source II, which is responsible for the formation of the cosmic ray spectrum in the range above the kink ($\epsilon_k\gtrsim 10^{15}$ eV), it is useful to consider its required properties. Above all, it is logical to inquire whether the spectrum is generically related to source I, which forms the cosmic ray spectrum at $\epsilon_k\lesssim 10^{15}$ eV. Indeed, a continuous cosmic ray spectrum formed without any break in summation of the spectra generated by the two sources can

be obtained only when several conditions are met. The maximum energy $\epsilon_{\max}$ in the source I spectrum must coincide with the minimum cosmic ray energy in the source II spectrum, and the amplitudes of both spectra at the point $\epsilon_{\max}$ must be identical. Inasmuch as the observed cosmic ray spectrum has one singularity, a kink in the spectrum at $\epsilon_k \approx 3 \times 10^{15}$ eV, it is desirable that the condition $\epsilon_{\max} \approx 3 \times 10^{15}$ eV be satisfied. If we assume that the two cosmic ray sources are independent, it seems highly improbable that the stated conditions will be satisfied. A more justifiable notion is the assumption that sources I and II are generically related, the second one serving as a preacceleration mechanism,[47] which entrains the highest-energy cosmic ray particles generated by the first source. The matching of the cosmic ray spectra is not a problem in this case. Estimates show that the preacceleration of cosmic rays can take place when they interact with the shock wave in the wind of young pulsars.[47]

If a pure power-law spectrum is formed during preacceleration, it is most naturally assumed that the total cosmic ray spectrum can be reproduced by extending the spectrum of cosmic rays from supernovae in the form of a power law $J_A(\epsilon_k) \propto \epsilon_k^{-3.1}$, beginning with the point $\epsilon_k = \epsilon_*$ at which the local exponent of the spectrum formed in supernova remnants $\gamma = -d \ln J_A / d \ln \epsilon_k$ is equal to 3.1. It is readily confirmed that the cosmic ray spectrum constructed in this way in the range $\epsilon_k > 10^{15}$ eV has a considerably lower amplitude than the observed spectrum. To achieve agreement with experiment, it is necessary that the maximum energy of cosmic rays accelerated in supernova remnants be approximately four times the energy corresponding to the calculated level shown in Fig. 2. To visualize this relation, we have calculated the cosmic ray spectrum for the hot phase of the interstellar medium with a magnetic field $B_0 = 12\ \mu$G and an injection rate $\eta = 5 \times 10^{-4}$. The cosmic ray spectrum extended into the high-energy region by the above-indicated technique is also shown in Fig. 2. It is evident from the figure that the calculations reproduce the measured cosmic ray spectrum well both before and after the kink. The required magnetic field $B_0 \approx 10\ \mu$G can be achieved for type Ib and type II supernovae, the acceleration of cosmic rays taking place in a medium significantly modified by the presupernova wind.[48]

Even though the procedure for plotting the cosmic ray spectrum in the range $\epsilon_k \gtrsim 10^{15}$ eV does not have a reliable physical foundation, it is useful in that it predicts not only the profile of the spectrum of cosmic rays, but also their composition in the indicated energy range. The cosmic ray composition is in fact particularly sensitive to the specific matching of the cosmic ray spectra from the two sources.

Figure 4 shows the results of calculations of the average atomic number of cosmic rays

$$\langle A \rangle = \sum J_A(\epsilon_k) A \Big/ \sum J_A(\epsilon_k) \tag{34}$$

as a function of the kinetic energy ϵ_k. The summation in Eq. (34) is carried out over the elements shown in Fig. 2.

It is evident from Fig. 4 that the quantity $\langle A \rangle$ increases rapidly in the range $\epsilon_k \lesssim 10^{12}$ eV up to $\langle A \rangle \simeq 10$ and then remains constant up to $\epsilon_k \sim 10^{15}$ eV. The rise of $\langle A \rangle$ in the range $\epsilon_k \lesssim 10^{12}$ eV is mainly attributable to the dependence

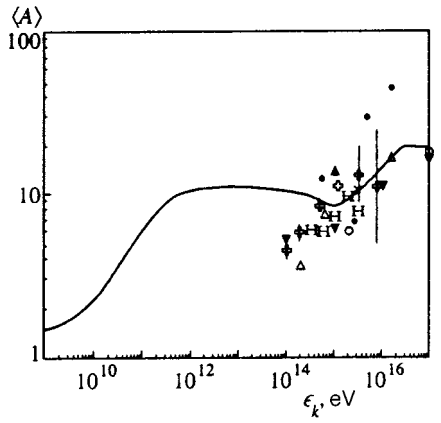

FIG. 4. Average atomic number of cosmic rays as a function of the kinetic energy. The experimental values are taken from Ref. 3, and the calculations correspond to the dot–dash curve in Fig. 3.

of the loss factor and the modulating factor on the hardness of the particles. Both factors increase the hardness of the spectrum in the range $\epsilon_k \lesssim Z$ GeV. This is why $\langle A \rangle$ increases as the energy is increased to $\epsilon_k \sim 10^{12}$ eV, at which both factors become relatively insignificant for iron nuclei.

The behavior of $\langle A \rangle$ in the range $\epsilon_k = 10^{15} - 10^{16}$ eV, increasing from $\langle A \rangle \approx 10$ to $\langle A \rangle \approx 20$, is entirely attributable to the dependence of the maximum cosmic ray energy on the nuclear charge $\epsilon_{\max} \propto Z$. The results of numerous indirect measurements,[3] shown in Fig. 4, are fully consistent with the predictions of the theory.

In the energy range $10^{14} - 10^{16}$ eV features can appear as a result of the contribution of local supernova remnants (i.e., the ones closest to the solar system).[49] Since the experimental evidence obtained in this regard indicates a critical ratio,[3] it could be helpful to examine this possibility theoretically. Accordingly, we write the average spectrum (differential number density with respect to the kinetic energy) of cosmic rays in the Galaxy in the form

$$n_c^G = \frac{e_c(\gamma - 2)}{(mc^2)^2} \left(\frac{\epsilon_k}{mc^2} \right)^{-\gamma}, \tag{35}$$

where $e_c \approx 1$ eV/cm^3 is the energy density of cosmic rays in the Galaxy, and $\gamma = 2.75$.

Owing to the hardness of the spectrum of cosmic rays accelerated in supernova remnants, their energy content is concentrated mainly in the relativistic part of the spectrum. Their density therefore obeys the approximate equation

$$n_c = \frac{\delta E_{\rm sn}}{V (mc^2)^2 \ln(\epsilon_{\max}/mc^2)} \left(\frac{\epsilon_k}{mc^2} \right)^{-2}, \tag{36}$$

in which $\delta E_{\rm sn}/V$ is the energy density of cosmic rays in the supernova remnants, δ is the fraction of the explosion energy transferred into the acceleration process in the cosmic rays, and $V = 4\pi R_c^3/3$ is the volume occupied by relativistic cosmic rays. In the active period of evolution of supernova remnants, $\Delta t \lesssim 10^4$ yr, when the bulk of the cosmic rays are generated, they are all concentrated in the volume occupied by the shock wave, implying that $R_c = R_s$, where R_s is the radius of the shock wave. In later phases of evolution, after the shock wave has weakened and ceases to accelerate cosmic

rays efficiently, the radius R_c of the volume occupied by them increases with time more rapidly than the shock radius R_s; this so-called runaway phenomenon first begins at the highest energies and then extends to lower energies as the shock wave weakens.[5,8]

The contribution of a solitary supernova remnant with a hard cosmic ray spectrum (36) superposed on the substantially softer galactic spectrum (35) is most noticeable in the vicinity of the maximum energies, $\epsilon_k \sim \epsilon_{\max}$. It is readily confirmed that the critical radius $R_c = R_*$, which is determined from the condition $n_c(\epsilon_{\max}) = n_c^G(\epsilon_{\max})$, is given by the expression

$$R_* = \left[\frac{3\,\delta E_{\rm sn}}{4\pi e_c(\gamma-2)\ln(\epsilon_{\max}/mc^2)} \left(\frac{\epsilon_{\max}}{mc^2} \right)^{\gamma-2} \right]^{1/3}. \qquad (37)$$

Substituting $\gamma = 2.75$, $E_{\rm sn} = 10^{51}$ erg, and $\delta = 0.3$ into this expression, we obtain $R_* \approx 1$ kps for an energy $\epsilon_{\max} = 10^5\, mc^2$. This means that at a definite time in the evolution (expansion) of a typical supernova remnant located at a distance $d \leqslant R_* = 1$ kps from the observer it creates an excess density of cosmic rays with energies $\epsilon_k \sim 10^{14}$ eV, such that $\Delta n_c \geqslant n_c^G$.

We now estimate the probability that an observer situated in the galactic disk will ''see'' this excess of the cosmic ray intensity. Inasmuch as supernovae in the Galaxy occupy a cylindrical region of radius $R \approx 10$ kps and height $h \approx 100$ ps, the probability that an individual supernova will be situated at a distance $d > R_*$ is $1-q$, where $q = R_*^2/R^2 = 10^{-2}$. The probability that all $N_{\rm sn}$ coexisting supernovae, whose growth is characterized by the condition $R_c(\epsilon_{\max}) = R_*$, will exist beyond the distance $d = 1$ kps is given by the expression

$$P = (1-q)^{N_{\rm sn}} \approx \exp(-N_{\rm sn} q).$$

The number of supernovae $N_{\rm sn} = \nu T$ is dictated by the frequency $\nu \approx 1/30\ {\rm yr}^{-1}$ of supernova flares in the Galaxy and their lifetime T. The growth of the cloud of cosmic rays in the stage $R_c > R_s$ is determined by their diffusion coefficient, because in this stage the region occupied by cosmic rays, $R_c \approx \sqrt{\kappa T}$, expands as a consequence of their diffusive propagation. In this stage the cosmic ray diffusion coefficient κ is a function not only of the energy, but also of time; κ increases with time, tending to the galactic average, because the level of turbulence generated by cosmic ray particles decreases as a result of the decrease in their energy density. Consequently, the determination of the increase in T reduces to the solution of the nonlinear problem of self-consistent expansion of the cosmic ray cloud. Owing to the complexity of this problem, it is advisable to consider the minimum estimate $T = R_c/c$. It is seen at once to yield $P \approx 1/e$. This means that with a probability $Q = 1-P$ close to unity the solar system must be situated in a cloud of cosmic rays produced by a local supernova, creating at least a twofold excess of the intensity of cosmic rays with energies $\sim 10^{14}$ eV above the galactic average. Since the observed cosmic ray spectrum does not exhibit any distinct hardening at these energies, it is logical to infer that the estimate of the probability Q derived above is far too high. The only one of the stated assumptions that can actually be violated is the sphericity of the expanding cosmic ray cloud. It is therefore justified to assume that in reality the cloud becomes highly asymmetric in the later stages, so that its radius in the galactic plane $R_\parallel$ is much smaller than the radius $R_\perp$ in the perpendicular direction. We therefore conclude that in the expansion stages of the cloud of cosmic rays, when their energy density is still appreciable, $e_c \gtrsim e_G$, the radius of the cloud $R_\perp$ in the direction perpendicular to the galactic plane can be much greater than 1 kps. Here, at least for energies $\epsilon_k \sim \epsilon_{\max}$, the average cosmic ray spectrum in the Galaxy $n_c^G(\epsilon_k)$ forms in such a way that cosmic rays produced by a solitary supernova remnant initially penetrate the region of the galactic halo and only then intermingle in the Galaxy with cosmic rays from other supernova remnants. The physical cause of this scenario is the cosmic ray pressure. Since the investigated phases of evolution of the cosmic ray cloud correspond to the condition $e_c \gtrsim e_G$, their pressure gradient is capable of generating in the interstellar medium a perturbation that expands predominantly in the direction of regions having the lowest density of interstellar matter, i.e., in the direction of the galactic halo. One cannot rule out the possibility of some of the cosmic rays breaking through into intergalactic space and inevitably escaping the Galaxy. The remaining cosmic rays become entangled in the magnetic field and contribute to the observed average cosmic ray spectrum in the Galaxy. Inasmuch as this effect extends into the main high-energy part of the cosmic ray spectrum, it can soften the observed spectrum, thereby offering a possible reconciliation of this discrepancy between the required and observed energy dependence of the residence time of cosmic rays in the Galaxy.

We note that the strong dependence of the volume occupied by cosmic ray particles on their energy in the later stages of evolution should not significantly influence the conclusions drawn above. It is essential only that the total number of cosmic rays produced in the active period of evolution of supernova remnants be described by the postulated energy dependence $N_c \propto \epsilon_k^{-2}$, along with the assumption that the cosmic ray energy not subsequently undergo any appreciable change. The fact that the expected spectrum of cosmic rays produced in supernova remnants is somewhat harder than the spectrum (36) merely reinforces the dynamical role of cosmic rays having the highest energies $\epsilon_k \sim \epsilon_{\max}$ and is conducive to actualization of this scenario.

If for the critical radius of the cosmic ray cloud in the galactic plane we take what we consider to be the minimum estimate, $R_\parallel \sim 100$ ps, the corresponding probability of an observer in the Galaxy ''seeing'' the effect from a local supernova remnant $P \sim 10^{-2}$ is still not negligible and can increase significantly if the maximum energy of cosmic rays accelerated in supernova remnants $\epsilon_{\max}$ is higher than the postulated value 10^{14} eV.

4. CONCLUSION

The foregoing analysis shows that cosmic rays accelerated in supernova remnants satisfactorily reproduce the observed profile of the energy spectrum of all the elements for which direct measurements have been made.

The theory is also capable of reproducing the observed dependence (increase) of the relative contents of elements in the composition of cosmic rays, $e(A)$, for a fixed energy per nucleon as a function of the atomic number A, i.e., the correct amplitude ratio of the spectra of various elements in their composition. Here the increase in $e(A)$ is partially attributable to the property of the nonlinear process of regular acceleration, as a result of which it more efficiently accelerates particles having a larger ratio A/Q of the mass and charge numbers. The quantitative reproduction of the dependence $e(A)$ requires that the injection mechanism also have the property of enriching with heavy elements: $e_{\rm inj} \propto A^{\alpha}$, $\alpha \approx 1$. The existing results of numerical simulation of collisionless, quasi-longitudinal shock waves bear witness to the existence of an injection mechanism with these properties.

A preliminary analysis has shown that the details of how and at what energy $\epsilon_{\max}$ the spectrum of cosmic rays produced in supernova remnants matches up with the spectrum formed in the range $\epsilon > \epsilon_{\max}$ by certain other type II sources significantly influence the resultant composition of the cosmic rays. For example, if we have $\epsilon_{\max} \sim 10^{15}$ eV the cosmic ray spectrum in the range $\epsilon > \epsilon_{\max}$ can be formed by the preacceleration of rays produced in supernova remnants. In the range $10^{15}-10^{16}$ eV the average atomic number of the cosmic rays $\langle A \rangle$ is expected to increase from $\langle A \rangle \simeq 10$ to $\langle A \rangle \simeq 20$ in this case, which does not contradict the existing experimental results. But if the maximum energy of cosmic rays accelerated in supernova remnants is $\epsilon_{\max} \sim 10^{14}$ eV or lower, we encounter a problem that is difficult to solve in our opinion: matching with the spectrum of source II. In addition, the expected ''weighting'' of cosmic rays shifts into the interval $10^{14}-10^{15}$ eV in this case, an event that is far less consistent with measurements (see Fig. 4).

Consequently, the maximum energy $\epsilon_{\max}$ to which cosmic rays can be accelerated in supernova remnants significantly affects their expected composition in the energy range $10^{14}-10^{16}$ eV. It is difficult to choose among the possible values of $\epsilon_{\max}$ solely on the basis of theoretical considerations. We have shown above that the maximum energy $\epsilon_{\max} \sim 10^{15}$ eV corresponding to the simplest solution of the problem of matching the spectra of cosmic rays from the two sources is attained for a magnetic field in the interstellar medium $B_0 = 12\,\mu$G, which is two to four times the value deemed to be typical of the unperturbed interstellar medium, $B_0 = 3-6\,\mu$G. However, we need to bear in mind several facts that could account for the indicated inconsistency. For example, it is justified to assume that the evolution of supernova remnants often takes place against the background of an already perturbed interstellar medium, in which it is entirely natural to postulate a somewhat elevated large-scale magnetic field B_0. The interstellar medium can then be perturbed both by previous supernova explosions and by the strong stellar winds of presupernovae, as is typical of type Ia and type II supernovae.[39] In our opinion, therefore, today the maximum energy $\epsilon_{\max} \sim 10^{15}$ eV should not be viewed as unjustifiably excessive.

According to the above estimates, the probability of a local supernova remnant making a measurable contribution to the observed cosmic ray spectrum in the energy range $\epsilon_k \gtrsim 10^{14}$ eV is not negligible, so that the recording of this phenomenon can yield direct information about the maximum energy of cosmic rays produced in supernova remnants.

The foregoing remark attests to the exceptional importance of new and reliable measurements of the spectrum of cosmic ray components in the range $10^{14}-10^{16}$ eV, as the results can be used to formulate a more justifiable conclusion as to the adequacy of the theory of regular acceleration and to ascertain the maximum energy of cosmic rays produced in supernova remnants.

In closing, the authors are pleased to express their thanks to A. W. Wolfendale, A. D. Erlykin, and G. F. Krymskiĭ for valuable discussions of the subjects discussed in the paper.

This work has received financial support from the Russian Fund for Fundamental Research (Grant No. 97-02-16132).

*)E-mail: berezho@sci.yakutia.ru
†)E-mail: ksenofon@sci.yakutia.ru

[1] V. S. Berezinskiĭ, S. V. Bulanov, V. L. Ginzburg *et al.*, *Astrophysics of Cosmic Rays*, North–Holland, Amsterdam (1990) [Russian orig., Nauka, Moscow (1984), p. 1].
[2] T. Shibata, in *Proceedings of the 24th International Cosmic Ray Conference: Invited, Rapporteur, and Highlight Papers*, Rome (1995), p. 713.
[3] A. Watson, in *Proceedings of the 25th International Cosmic Ray Conference: Invited, Rapporteur, and Highlight Papers*, Durban, South Africa (1997), p. 257.
[4] L. O'C. Drury, Rep. Prog. Phys. **46**, 973 (1983).
[5] E. G. Berezhko and G. F. Krymskiĭ, Usp. Fiz. Nauk **154**, 49 (1988) [Sov. Phys. Usp. **31**, 27 (1988)].
[6] E. G. Berezhko, V. K. Yelshin, and L. T. Ksenofontov, Astropart. Phys. **2**, 215 (1994).
[7] E. G. Berezhko, V. K. Yelshin, and L. T. Ksenofontov, Nucl. Phys. **2**, 215 (1994).
[8] E. G. Berezhko, V. K. Elshin, and L. T. Ksenofontov, Zh. Éksp. Teor. Fiz. **109**, 3 (1996) [JETP **82**, 1 (1996)].
[9] E. G. Berezhko and H. J. Völk, Astropart. Phys. **7**, 183 (1997).
[10] G. F. Krymskiĭ, Dokl. Akad. Nauk SSSR **234**, 1306 (1977) [Sov. Phys. Dokl. **22**, 327 (1977)].
[11] W. I. Axford, E. Leer, and G. Skadron, in *Proceedings of the 15th International Cosmic Ray Conference*, Plovdiv, Bulgaria (1977), Vol. 11, p. 132.
[12] R. D. Blandford and J. P. Ostriker, Astrophys. J. **221**, L29 (1978).
[13] G. F. Krymskiĭ, *Geomagn. Aeron.* **4**, 977 (1964).
[14] E. N. Parker, Planet. Space Sci. **13**, 9 (1965).
[15] L. O'C. Drury, W. J. Markiewicz, and H. J. Völk, Astron. Astrophys. **225**, 179 (1989).
[16] E. G. Berezhko, G. F. Krymsky, and A. A. Turpanov, in *Proceedings of the 21th International Cosmic Ray Conference*, Adelaide, Australia (1990), Vol. 4, p. 101.
[17] H. Kang and T. W. Jones, Mon. Not. R. Astron. Soc. **249**, 439 (1991).
[18] L. O'C. Drury and H. J. Völk, Astrophys. J. **248**, 344 (1981).
[19] L. O'C. Drury, H. J. Völk, and E. G. Berezhko, Astron. Astrophys. **299**, 222 (1994).
[20] E. G. Berezhko, Astropart. Phys. **5**, 367 (1996).
[21] A. R. Bell, Mon. Not. R. Astron. Soc. **182**, 147 (1978).
[22] J. F. McKenzie and H. J. Völk, Astron. Astrophys. **116**, 191 (1982).
[23] E. A. Dorfi and L. O'C. Drury, in *Proceedings of the 19th International Cosmic Ray Conference*, La Jolla, California (1985), Vol. 3, p. 121.
[24] L. O'C. Drury, Adv. Space Res. **4**, 185 (1984).
[25] E. G. Berezhko, Pis'ma Astron. Zh. **12**, 842 (1986) [Sov. Astron. Lett. **12**, 352 (1986)].
[26] E. M. Jones, B. W. Smith, and W. C. Straka, Astrophys. J. **249**, 185 (1981).
[27] R. A. Chevalier, Astrophys. J. **259**, 302 (1982).
[28] R. A. Chevalier and E. P. Liang, Astrophys. J. **344**, 332 (1989).

[29] E. A. Dorfi, Astron. Astrophys. **234**, 419 (1990).
[30] G. G. Chernyĭ, Dokl. Akad. Nauk SSSR **112**, 113 (1957).
[31] E. A. Dorfi, Astron. Astrophys. **251**, 597 (1991).
[32] K. B. Quest, J. Geophys. Res. **93**, 9649 (1988).
[33] K. J. Trattner and M. Scholer, Ann. Geophys. (Germany) **9**, 774 (1993).
[34] Y. T. Gosling, J. R. Asbridge, S. J. Bame *et al.*, J. Geophys. Res. **86**, 547 (1981).
[35] K. J. Trattner, E. Möbius, M. Scholer *et al.*, J. Geophys. Res. **99**, 389 (1994).
[36] D. C. Ellison, F. C. Jones, and D. Eichler, J. Geophys. Res. **50**, 110 (1981).
[37] M. A. Malkov and H. J. Völk, Astron. Astrophys. **300**, 605 (1995).
[38] M. G. Baring, K. W. Ogilvie, D. C. Ellison, and R. J. Forsyth, Astrophys. J. **476**, 889 (1997).
[39] T. A. Losinskaya, in *Proceedings of the 22th International Cosmic Ray Conference*, Dublin (1991), Vol. 5, p. 123.
[40] S. A. Kaplan and S. B. Pinkel'ner, *Physics of the Interstellar Medium* [in Russian], Nauka, Moscow (1979).
[41] L. I. Dorman, *Cosmic Rays: Variations and Space Explorations*, North-Holland, Amsterdam (1974) [Russian orig., Izd. AN SSSR, Moscow (1963)].
[42] N. L. Grigorov, Kosm. Issled. **33**, 339 (1995).
[43] J. P. Meyer, L. O'C. Drury, and D. C. Ellison, Astrophys. J. **487**, 182 (1997).
[44] E. G. Berezhko and D. C. Ellison, submitted to Astrophys. J. (1999).
[45] E. G. Berezhko, G. F. Krymsky, V. K. Yelshin, and L. T. Ksenofontov, in *Proceedings of the 25th International Cosmic Ray Conference*, Durban, South Africa (1997), Vol. 4, p. 417.
[46] D. C. Ellison, L. O'C. Drury, and J. P. Meyer, Astrophys. J. **487**, 197 (1997).
[47] E. G. Berezhko, Pis'ma Astron. Zh. **20**, 93 (1994) [Sov. Astron. Lett. **20**, 75 (1994)].
[48] E. G. Berezhko and H. J. Völk, submitted to Astropart. Phys. (1999).
[49] A. D. Erlykin and A. W. Wolfendale, Astropart. Phys. **7**, 1 (1997).

Translated by James S. Wood

Interpolation of experimental data without a theoretical model

A. D. Bukin*)

G. I. Budker Institute of Nuclear Physics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 26 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 760–776 (September 1999)

A new approach is proposed for the interpolation of experimental data when a theoretical model is unavailable. The method is based on the minimization of a modified likelihood function incorporating a generalized smoothness test of the theoretical curve. The conditions for applicability of the method are a sufficiently accurate estimate of the statistical errors of the experimental data (assuming that the results fit a Gaussian distribution in accordance with the resulting rms deviations) and smoothness of the theoretical curve. A FORTRAN 77 program for the interpolation of experimental data has been written to implement the proposed algorithm. The computing time and roundoff error are determined as functions of the number of experimental points. © *1999 American Institute of Physics.* [S1063-7761(99)00209-7]

1. INTRODUCTION

The interpolation of experimental data poses one of the most important problems in experimental physics. In the majority of cases a theoretical model (possibly more than one) is available for the accumulated data, permitting known methods to be used for optimal estimation of the model parameters (the most popular methods are least squares and maximum likelihood; see, e.g., Refs. 1 and 2). There are times, however, when it is impossible to use a theoretical model for data interpolation, but the theoretical curve describing the experimental data is positively known to be smooth. Either polynomials or polynomial splines are customarily used in such cases. The objective of the present study is to develop a generalized algorithm for the interpolation and smoothing of experimental data when statistical errors are present at all measured points and a theoretical model does not exist. A recent paper on this subject[3] sets forth the basic notion of choosing a linear combination of cubic B splines with a uniform computational grid as the smoothing function and of inferring the statistical behavior of the deviations of the experimental points from the theoretical curve by decreasing the number of spline nodes I_{opt} until satisfactory agreement is achieved between theory and experiment by the χ^2 test. In this paper a generalized smoothness test is proposed for the approximation of experimental data, whereby the degree of smoothness of the curve is varied until satisfactory agreement is achieved between theory and experiment.

2. CHOOSING AN APPROXIMATING FUNCTION

Let it be required to find a smooth function $f(x)$ that passes through n experimental values y_i with errors σ_i (assumed to have a Gaussian distribution) at points x_i. The standard approach to such problems is to minimize the logarithmic likelihood function (or χ^2)

$$L=\sum_{i=1}^{n}\frac{[f(x_i)-y_i]^2}{\sigma_i^2}. \tag{1}$$

In contrast with the standard approach, there are still no model parameters by which to minimize the function L. Below, we must somehow formulate a smoothness test. To do so, we propose minimizing the functional

$$I_S=\int_{-\infty}^{\infty}[f''(x)]^2dx. \tag{2}$$

To unify the smoothness test and achieve agreement with the experimental data, we modify the likelihood function as follows:

$$\tilde{L}=L+wI_S, \tag{3}$$

where w is a positive weighting factor. This technique of consolidating the likelihood function and the smoothness test into a single objective function has in fact been used[4] in solving the problem of smoothing experimental data by second-order and third-order splines. In Ref. 4, however, no basis is given to explain why polynomial splines best approximate the experimental data. Nor in Ref. 4 are any recommendations given for choosing the coordinates of the spline nodes or the weighting factor w (the purely mathematical problem is discussed in the book). In the proposed algorithm the weighting factor is increased until L attains the value $L_0=n-2$. This value is in fact equal to the average value of χ^2 in fitting the experimental values to a theoretical curve with n degrees of freedom, subject to the condition that the overall level and slope are free parameters (although a somewhat different level can also be set).

It is obvious that outside the range of experimental points x over which the required function varies, there are no restrictions on the value of the function and, hence, no reason for any deviation from a straight line; we can therefore assume that the required function has the minimum possible value $[f''(x)]^2$:

$$f''(x)=0 \quad \text{for} \quad x<x_1 \quad \text{or} \quad x>x_n. \tag{4}$$

We attempt to find a solution by methods of the variational calculus.[5] Following the standard variational calculus approach, we introduce a small arbitrary function $\delta f(x)$ and determine the variation $\tilde{L}$:

$$\delta L=2\sum_{i=1}^{n}\frac{f(x_i)-y_i}{\sigma_i^2}\delta f(x_i)$$

$$=2\int_{-\infty}^{\infty}\left[\sum_{i=1}^{n}\frac{f(x)-y_i}{\sigma_i^2}\delta f(x)\,\delta(x-x_i)\right]dx, \tag{5}$$

where

$$\delta(x)=\begin{cases}0 & \text{for } x\neq 0,\\ \infty & \text{for } x=0,\end{cases} \qquad \int_{-\infty}^{\infty}\delta(x)dx=1$$

and

$$\delta I_S=2\int_{-\infty}^{\infty}f''(x)\,\delta f''(x)dx=2\int_{-\infty}^{\infty}f''''(x)\,\delta f(x)dx. \tag{6}$$

Here we have made use of the fact that

$$f''(x)\to 0, \quad f'''(x)\to 0 \quad \text{as} \quad x\to\pm\infty. \tag{7}$$

The variation of the functional $\tilde{L}$ must now be set equal to zero:

$$\delta\tilde{L}=2\int_{-\infty}^{\infty}\delta f(x)dx\left[f''''(x)+\sum_{i=0}^{n}\frac{f(x)-y_i}{\sigma_i^2}\delta(x-x_i)\right]=0. \tag{8}$$

Since the function $\delta f(x)$ is arbitrary, Eq. (8) is valid only if

$$f''''(x)+\sum_{i=0}^{n}\frac{f(x)-y_i}{\sigma_i^2}\delta(x-x_i)=0. \tag{9}$$

This equation implies that in the intervals between the experimental points x_i and x_{i+1} the derivative $f''''(x)$ is identically zero, which is one definition of an arbitrary third-degree polynomial, and one or more of the leading derivatives can have discontinuities at the experimental measurement points. This fact uniquely determines the type of approximating function $f(x)$: a cubic spline of deficiency 1, 2, or 3 (in the classification of splines in Ref. 4 the deficiency of a spline is the number of leading derivatives that have discontinuities at the nodes of the spline). The coordinates of the spline nodes are also uniquely determined: They are the coordinates of the experimental points x_i.

We now consider the case of splines of deficiency 3. This solution obviously represents a broken line passing through the experimental values y_i. The functional $\tilde{L}$ is equal to zero in this case. The solution is mathematically possible, but is of no practical interest, as only the cases of deficiencies 1 and 2 will be discussed below.

We choose a spline parametrization in terms of the values of the spline S_i and the values of the derivative of the spline D_i at the nodes x_i. The value of the spline at an intermediate point is then

$$f(x)=S_i p_1(x;x_i,x_{i+1})+S_{i+1}p_2(x;x_i,x_{i+1})$$
$$+D_i q_1(x;x_i,x_{i+1})+D_{i+1}q_2(x;x_i,x_{i+1}),$$
$$x_i\leqslant x\leqslant x_{i+1}, \tag{10}$$

where

$$p_1(x;x_1,x_2)=\frac{(x-x_2)^2(2x+x_2-3x_1)}{(x_2-x_1)^3},$$

$$q_1(x;x_1,x_2)=\frac{(x-x_2)^2(x-x_1)}{(x_2-x_1)^2},$$

$$p_2(x;x_1,x_2)=\frac{(x-x_1)^2(3x_2-2x-x_1)}{(x_2-x_1)^3},$$

$$q_2(x;x_1,x_2)=\frac{(x-x_1)^2(x-x_2)}{(x_2-x_1)^2}. \tag{11}$$

Indeed, any third-degree polynomial can be written as a linear combination of polynomials p_1,p_2,q_1,q_2. On the other hand, the indicated standard polynomials have readily verifiable properties:

$$p_1(x_2;x_1,x_2)=p_1'(x_1;x_1,x_2)=p_1'(x_2;x_1,x_2)=0,$$

$$p_2(x_1;x_1,x_2)=p_2'(x_1;x_1,x_2)=p_2'(x_2;x_1,x_2)=0,$$

$$q_1(x_1;x_1,x_2)=q_1(x_2;x_1,x_2)=q_1'(x_2;x_1,x_2)=0,$$

$$q_2(x_1;x_1,x_2)=q_2(x_2;x_1,x_2)=q_2'(x_1;x_1,x_2)=0,$$

$$p_1(x_1;x_1,x_2)=p_2(x_2;x_1,x_2)=q_1'(x_1;x_1,x_2)$$
$$=q_2'(x_2;x_1,x_2)=1, \tag{12}$$

so that they can be used in the form (10) for the parametrization of an arbitrary cubic spline of deficiency 2. This kind of parametrization is particularly well-suited to splines in a multidimensional space (which the present author has used, e.g., in Ref. 6). The second derivatives to the right and to the left of x_i are expressed in terms of the coefficients S_i and D_i:

$$f''(x_i-0)=\frac{2(D_{i-1}+2D_i)}{x_i-x_{i-1}}+\frac{6(S_{i-1}-S_i)}{(x_i-x_{i-1})^2},$$

$$f''(x_i+0)=-\frac{2(D_{i+1}+2D_i)}{x_{i+1}-x_i}+\frac{6(S_{i+1}-S_i)}{(x_{i+1}-x_i)^2}. \tag{13}$$

Consequently, for cubic splines of deficiency 1 (the most common type) the following conditions for continuity of the second derivative must be satisfied for the coefficients D_i:

$$\left.\frac{d^2f}{dx^2}\right|_{x=x_i}=C_i=\frac{2(D_{i-1}+2D_i)}{x_i-x_{i-1}}+\frac{6(S_{i-1}-S_i)}{(x_i-x_{i-1})^2}$$

$$=-\frac{2(D_{i+1}+2D_i)}{x_{i+1}-x_i}+\frac{6(S_{i+1}-S_i)}{(x_{i+1}-x_i)^2}. \tag{14}$$

The set of coefficients S_i,C_i could be used for the determination of splines of deficiency 1, whereupon the coefficients D_i (and the condition for continuity of the first derivative) could be written

$$\left.\frac{df}{dx}\right|_{x=x_i} = D_i = \frac{S_{i+1}-S_i}{x_{i+1}-x_i} - \frac{x_{i+1}-x_i}{6}(2C_i + C_{i+1})$$

$$= \frac{S_i - S_{i-1}}{x_i - x_{i-1}} + \frac{x_i - x_{i-1}}{6}(2C_i + C_{i-1}). \quad (15)$$

For brevity, from now on we denote the distances between consecutive nodes by

$$h_i = x_{i+1} - x_i. \quad (16)$$

3. MINIMIZATION OF THE MEAN-SQUARE SMOOTHNESS PARAMETER I_S

Looking at the problem of minimizing the objective function (3) for splines of deficiency 2, we note that L does not depend on the parameter D_i. In the first stage, therefore, we can minimize the function $\tilde{L}$ with respect to D_i (in fact minimize I_S). Invoking the expression for a spline in terms of S_i and D_i, we readily obtain an equation for I_S:

$$I_S = \sum_{i=1}^{n-1} \left\{ \frac{1}{h_i}[D_{i+1} - D_i]^2 + \frac{3}{h_i^3} \times [h_i(D_{i+1} + D_i) + 2(S_i - S_{i+1})]^2 \right\}. \quad (17)$$

If we write the system of equations specifying the condition of the minimum of I_S with respect to D_i, we find that it is equivalent to the condition of continuity of the second derivative at all intermediate nodes and vanishing of the second derivative at the extreme nodes. We find that a cubic spline of deficiency 2 ''shuns'' the additional degree of freedom and reverts to a spline of deficiency 1 for the minimization of I_S. In the ensuing discussion, therefore, we shall not mention this characteristic of the spline every time, bearing in mind that only cubic splines of deficiency 1 are considered. For a spline of deficiency 1 the function I_S attains a minimum for $C_1 = C_n = 0$. This result is well known and is used to draw a spline through points with given values of S_i (Ref. 4). We have thus exhausted the possibilities of minimizing the ''smoothness parameter'' I_S separately and must now minimize the entire function $\tilde{L}$ with respect to the remaining n free parameters.

Considering that one of our practical goals is to be able to approximate large arrays (at least in the dozens) of experimental points, we have no choice but to complicate the investigation by using large matrices. One of the most effective techniques for obtaining almost diagonal matrices and, hence, significantly abating the effects of roundoff errors is to represent the spline by a linear combination of B splines. This technique has been employed to great advantage in previous work.[3]

4. REPRESENTATION OF A SPLINE BY A LINEAR COMBINATION OF *B*-SPLINES

A B spline is a special type of cubic spline.[4] A distinguishing attribute of B splines is that they are identically zero everywhere except in a few consecutive intervals between nodes of the given spline. By constructing a system of B splines that covers the entire grid of spline nodes it is possible to represent any spline on this grid by a linear combination of B splines.

Any cubic spline on a grid $x_1, \ldots, x_n$ can be represented by a linear combination of B splines:

$$f(x) = \sum_{i=0}^{n+1} \beta_i B_i(x), \quad (18)$$

where

$$B_i(x) \equiv 0 \quad \text{for} \quad x \leqslant x_{i-2} \quad \text{or} \quad x \geqslant x_{i+2},$$

$$B_i(x_{i-2}) = B_i'(x_{i-2}) = B_i''(x_{i-2}) = 0,$$

$$B_i(x_{i+2}) = B_i'(x_{i+2}) = B_i''(x_{i+2}) = 0.$$

The technical details of implementing this algorithm can be found in Ref. 7.

If all the parameters β_i are free, the second derivative of the spline defined in Eq. (18) can have an arbitrary value at the extreme nodes. On the other hand, to minimize I_S and ensure continuity of the second derivative, we must use splines for which $f''(x_1) = f''(x_n) = 0$. These conditions are easily satisfied by the proper choice of coefficients β_0 and β_{n+1}, which is equivalent to modifying the B splines near the extreme nodes, so that

$$\tilde{B}_i''(x_1) = \tilde{B}_i''(x_n) = 0.$$

Our required splines are now written as follows on the interval (x_1, x_n):

$$f(x) = \sum_{i=1}^{n} \beta_i \tilde{B}_i(x), \quad (19)$$

where β_i can take any values.

In this notation the likelihood function is written in the form

$$L = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\beta_i\beta_j - 2\sum_{i=1}^{n} P_i\beta_i + \sum_{i=1}^{n} \frac{y_i^2}{\sigma_i^2}, \quad (20)$$

where

$$A_{ij} = \sum_{k=1}^{n} \frac{\tilde{B}_i(x_k)\tilde{B}_j(x_k)}{\sigma_k^2}, \quad P_i = \sum_{k=1}^{n} \frac{y_k \tilde{B}_i(x_k)}{\sigma_k^2}. \quad (21)$$

If we define the matrix $\mathbf{B}$ and the vector $\mathbf{Y}$ by the equations

$$B_{ij} = \tilde{B}_i(x_j), \quad Y_i = y_i/\sigma_i^2, \quad (22)$$

we can rewrite the expressions for L and P in the form

$$L = \beta^T \mathbf{A}\beta - 2\mathbf{P}^T\beta + \sum_{i=1}^{n} \frac{y_i^2}{\sigma_i^2}, \quad \mathbf{P} = \mathbf{B}\mathbf{Y}, \quad (23)$$

where the superscript T denotes transposition.

5. ALGORITHM FOR CHOOSING THE WEIGHTING FACTOR

The modified likelihood function can now be written as

$$\tilde{L} = \beta^T \mathbf{A}\beta - 2\mathbf{P}^T\beta + \sum_{i=1}^{n} \frac{y_i^2}{\sigma_i^2} + w\beta^T\mathbf{R}\beta, \quad (24)$$

where $\mathbf{R}$ is a symmetric matrix,

$$R_{ij}=\int_{x_1}^{x_n} B_i''(x)B_j''(x)dx. \tag{25}$$

The algorithm for simultaneously reducing two symmetric matrices to diagonal form by a linear transformation is well known.[8,9] Following this algorithm, we can obtain both likelihood functions in the form (see Ref. 7 for details)

$$L=\mathbf{z}^T\mathbf{z}-2\mathbf{E}^T\mathbf{z}+\sum_{i=1}^{n}\frac{y_i^2}{\sigma_i^2},\quad \tilde{L}=L+w\cdot\mathbf{z}^T\mathbf{\Lambda}\mathbf{z},$$

$$\Lambda_{ij}=\lambda_i\delta_{ij}. \tag{26}$$

The system of linear equations for finding the minimum of $\tilde{L}$ can be separated into independent equations

$$\frac{1}{2}\frac{\partial\tilde{L}}{\partial z_i}=z_i-E_i+w\lambda_i z_i=0, \tag{27}$$

which have the solutions

$$z_i=\frac{E_i}{1+\lambda_i w},\quad i=1,\ldots,n. \tag{28}$$

We substitute the resulting solution into the equation for L and obtain L as a function of w:

$$L=\sum_{i=1}^{n}\frac{y_i^2}{\sigma_i^2}-\sum_{i=1}^{n}\frac{E_i^2(1+2\lambda_i w)}{(1+\lambda_i w)^2}. \tag{29}$$

In the limit $w\to 0$, obviously, the spline must pass through the experimental points, and we must have $L\to 0$, i.e.,

$$\sum_{i=1}^{n}\frac{y_i^2}{\sigma_i^2}=\sum_{i=1}^{n}E_i^2. \tag{30}$$

This identity, incidentally, can be used to test the validity of the equations in application to the implementation of the algorithm in computer programs.

Bearing the latter consideration in mind, we can reduce the equation for L to the form

$$L=\sum_{i=1}^{n}\frac{(E_i\lambda_i w)^2}{(1+\lambda_i w)^2}. \tag{31}$$

Each term in this sum increases monotonically as w increases, so that L is also a monotonically increasing function of w. The equation

$$L=L_0=n-2 \tag{32}$$

can be solved for w by any numerical method such as, for example, the bisection algorithm.[8]

It can happen that L remains smaller than L_0 even when the weighting factor is increased to infinity. This condition must be checked before starting the iteration cycle. The maximum possible value of L is obtained when the experimental data are interpolated by a straight line, with

$$\max L=L_{\max}=\sum_{i=1}^{n}\frac{(y_i-y_0)^2-a_1^2(x_i-x_0)^2}{\sigma_i^2}, \tag{33}$$

where

$$a_0=\sum_{i=1}^{n}\frac{1}{\sigma_i^2},\quad x_0=\frac{1}{a_0}\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2},$$

$$y_0=\frac{1}{a_0}\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2},\quad a_2=\sum_{i=1}^{n}\frac{(x_i-x_0)^2}{\sigma_i^2},$$

$$a_1=\frac{1}{a_2}\sum_{i=1}^{n}\frac{(x_i-x_0)(y_i-y_0)}{\sigma_i^2}. \tag{34}$$

Once the weighting factor w and the corresponding coefficients β_i have been chosen, the complete set of coefficients S_i and D_i can be obtained from the equations

$$\mathbf{S}=\mathbf{B}^T\beta,\quad D_i=\sum_{j=1}^{n}\beta_j\tilde{B}_j'(x_i). \tag{35}$$

When the spline degenerates into a straight line, the expressions for S_i and D_i are simplified:

$$S_i=y_0+a_1(x_i-x_0),\quad D_i=a_1. \tag{36}$$

6. ESTIMATION OF STATISTICAL ERROR

In estimating the statistical error, we shall assume that the experimental measurements at different points are statistically independent. In principle, correlations can be taken into account in the linear approximation if the covariance matrix is known for the experimental data.

At an arbitrary point $x\in(x_i,x_{i+1})$, according to Eq. (10), the following equation can be written for the variance of the spline:

$$\begin{aligned}\sigma^2(x)=&p_1^2(\langle S_i^2\rangle-\langle S_i\rangle^2)+p_2^2(\langle S_{i+1}^2\rangle-\langle S_{i+1}\rangle^2)\\&+q_1^2(\langle D_i^2\rangle-\langle D_i\rangle^2)+q_2^2(\langle D_{i+1}^2\rangle-\langle D_{i+1}\rangle^2)\\&+2p_1p_2(\langle S_iS_{i+1}\rangle-\langle S_i\rangle\langle S_{i+1}\rangle)\\&+2q_1q_2(\langle D_iD_{i+1}\rangle-\langle D_i\rangle\langle D_{i+1}\rangle)\\&+2p_1q_1(\langle S_iD_i\rangle-\langle S_i\rangle\langle D_i\rangle)+2p_1q_2(\langle S_iD_{i+1}\rangle\\&-\langle S_i\rangle\langle D_{i+1}\rangle)+2q_1p_2(\langle D_iS_{i+1}\rangle-\langle D_i\rangle\langle S_{i+1}\rangle)\\&+2q_2p_2(\langle D_{i+1}S_{i+1}\rangle-\langle D_{i+1}\rangle\langle S_{i+1}\rangle),\end{aligned} \tag{37}$$

where for brevity the standard polynomials from Eqs. (11) are denoted by

$$p_k=p_k(x;x_i,x_{i+1}),\quad q_k=q_k(x;x_i,x_{i+1}),\quad k=1,2.$$

For $x<x_1$ the spline goes over to the straight line $f(x)=S_1+D_1(x-x_1)$, so that

$$\begin{aligned}\sigma^2(x)|_{x<x_1}=&(\langle S_1^2\rangle-\langle S_1\rangle^2)+(x-x_1)^2(\langle D_1^2\rangle-\langle D_1\rangle^2)\\&+2(x-x_1)(\langle S_1D_1\rangle-\langle S_1\rangle\langle D_1\rangle).\end{aligned} \tag{38}$$

Analogously, for $x>x_n$

$$\begin{aligned}\sigma^2(x)|_{x>x_n}=&(\langle S_n^2\rangle-\langle S_n\rangle^2)+(x-x_n)^2(\langle D_n^2\rangle-\langle D_n\rangle^2)\\&+2(x-x_n)(\langle S_nD_n\rangle-\langle S_n\rangle\langle D_n\rangle).\end{aligned} \tag{39}$$

The exact calculation of the necessary correlation functions corresponding to the proposed algorithm poses a difficult

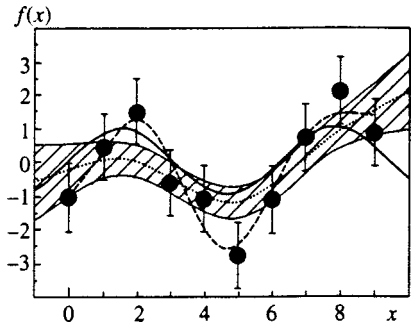

FIG. 1. Interpolation of points on a sine wave ($\sigma_i=1$). The solid curve corresponds to the ''exact'' theory, the dotted curve bordered by an error strip corresponds to the proposed approximation algorithm, and the dashed curve corresponds to interpolation by B splines.[3]

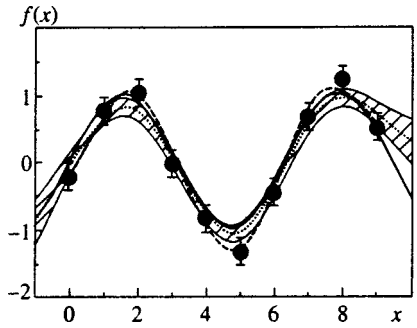

FIG. 2. Interpolation of points on a sine wave (very small errors). The solid curve corresponds to the ''exact'' theory, the dotted curve bordered by an error strip corresponds to the proposed approximation algorithm, and the dashed curve corresponds to interpolation by B splines.[3]

task. However, the statistical error can be adequately estimated in the approximation of a fixed weighting factor w. These correlation functions can be calculated during fitting of the experimental data and can be stored in an auxiliary file to be retrieved whenever needed for estimating the statistical curve-fitting error at an arbitrary point x. If the factor w is assumed to be fixed, the relations for S_i and D_i become linear in y_j, and the correlation functions are expressed in terms of matrix transformations used to diagonalize quadratic forms (see Ref. 7 for details).

7. TESTING THE OPERATION OF THE ALGORITHM IN EXAMPLES

To implement the above-described algorithm, a FITAB program has been written in FORTRAN 77 for the approximation of experimental data, along with corresponding VALSPL and ERRSPL programs for calculating the value of a spline and estimating the statistical error.

We now examine a few simple applications of the algorithm.

7.1. Sine wave with large experimental errors

We choose ten points situated around a sine wave of unit amplitude:

$$x_i=i-1 \quad \langle y_i\rangle=\sin x_i, \quad i=1,2,\ldots,10. \tag{40}$$

The points are dispersed relative to the sine wave according to a Gaussian law by means of an appropriate pseudorandom number generator:

$$\frac{dW}{dy_i}=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{(y_i-\langle y_i\rangle)^2}{2\sigma_i^2}\right]. \tag{41}$$

The solid curve in Fig. 1 represents the sine wave $y=\sin x$ (exact theory), and the ten points with error bars are obtained for $\sigma_i=1$.

The dashed curve represents results obtained by Anikeev *et al.*[3] (using the CSPLS1 program). Owing to the considerable latitude in choosing the CSPLS1 access parameters, it is necessary to fix certain principles for the selection of these parameters. We assume that the limits of approximation by B splines are $X_{\min}=x_1$ and $X_{\max}=x_n$ and that the number I_{opt} of B splines used for the approximation are chosen according to the authors' recommendations by the following algorithm:

1) $I_{\text{opt}}=2$.

2) For a given number of B splines I_{opt}, approximate the experimental quantities and calculate χ^2.

3) If $\chi^2>n_D=n-I_{\text{opt}}$ (the number of effective degrees of freedom) holds, increment once the number of B splines ($I_{\text{opt}}+1\to I_{\text{opt}}$) and go to step 2).

In this figure the dotted line bordered by an error strip represents the results obtained by our proposed algorithm.

To obtain quantitative criteria for assessing the quality of approximation, we make use of the fact that the exact theoretical curve is known here. We introduce two characteristics of the deviation of the approximating curve $f(x)$ from the true theoretical curve $f_{\text{exact}}(x)$: the maximum deviation of the absolute value

$$\Delta f_{\text{abs}}=\max_{x_1<x<x_n}|f(x)-f_{\text{exact}}(x)| \tag{42}$$

and the rms deviation

$$\Delta f_{\text{rms}}=\left\{\frac{1}{x_n-x_1}\int_{x_1}^{x_n}[f(x)-f_{\text{exact}}(x)]^2dx\right\}^{1/2}. \tag{43}$$

The results of approximation by the two compared algorithms, shown in Fig. 1, are characterized by the parameters

Smoothing algorithm	δf_{abs}	δf_{rms}
Ref. 3	1.60	0.79
This paper	1.09	0.57

The proposed algorithm has noticeably better characteristics in this case.

7.2. Sine wave with very small experimental errors

We now decrease the ''experimental'' errors by a factor of five ($\sigma_i=0.2$) and repeat the operation (Fig. 2). It is evident that the curvature of the interpolating curve has increased automatically, despite the invariance of the number of spline nodes. The approximation characteristics for this

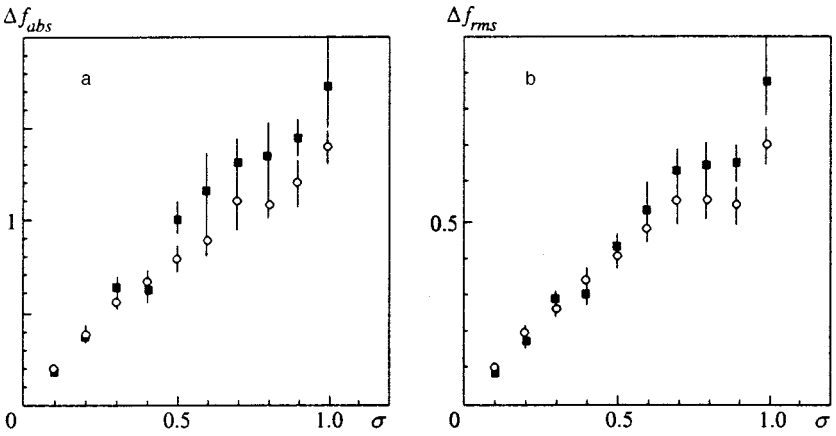

FIG. 3. Average values of the maximum $\overline{\Delta f_{\rm abs}}$ (a) and rms $\overline{\Delta f_{\rm rms}}$ (b) deviations of the approximating curve from the exact ''theoretical'' curve (a sine wave) for two interpolation algorithms: one using B splines (■) from Ref. 3 and the algorithm proposed in this paper (○).

case are as follows:

Smoothing algorithm	$\delta f_{\rm abs}$	$\delta f_{\rm rms}$
Ref. 3	0.35	0.18
This paper	0.28	0.12

Even though our above-defined ''quality of approximation'' parameters indicate that the proposed algorithm for this example provides a closer approximation to the true curve, it is still impossible to draw any conclusion on the basis of one statistical sample. Figure 3 shows the quality of approximation as a function of the errors at each experimental point. For each value of the experimental error we have approximated several sets of experimental points obtained by means of a pseudorandom number generator, and the graph shows the value of $\Delta f_{\rm abs}$ or $\Delta f_{\rm rms}$ averaged over several samples.

7.3. Broken line

We now perform an analogous operation with a broken line:

$$x_i = i-1, \quad \langle y_i \rangle = 0.3 \cdot |x_i - 5|, \quad i = 1,2,\ldots,10. \tag{44}$$

We carry out two series of fittings in precisely the same way, once with $\sigma_i = 1$ (Fig. 4) and once with $\sigma_i = 0.2$ (Fig. 5). The approximation characteristics for these cases are as follows:

Smoothing algorithm	$\Delta f_{\rm abs}$	$\Delta f_{\rm rms}$	$\Delta f_{\rm abs}$	$\Delta f_{\rm rms}$
	$\sigma = 1$		$\sigma = 0.2$	
Ref. 3	1.02	0.61	0.27	0.15
This paper	0.99	0.45	0.15	0.10

Again, for large errors, when the nonuniformity of the approximated quantity is less than or of the same order as the experimental errors, the given algorithm chooses a straight or almost-straight line for approximation, and for errors much smaller than the nonuniformity the approximation comes close to the theoretical curve.

As in the preceding example, the quality of approximation in the proposed algorithm is generally not much better than for interpolation by B splines (Fig. 6).

7.4. Gaussian peak against a flat background

Finally, we consider a more popular distribution for high-energy physics—a Gaussian peak against a flat background:

$$x_i = 9\,\frac{i-1}{n-1}, \quad i = 1,2,\ldots,n,$$

$$\langle y_i \rangle = 1 + 0.05 x_i + \exp[-2(x_i - 5)^2]. \tag{45}$$

Figure 7 shows the results of fitting such data for $n = 60$ and $\sigma_i = 0.2$. In this case the quality of approximation for the two algorithms has been evaluated as follows:

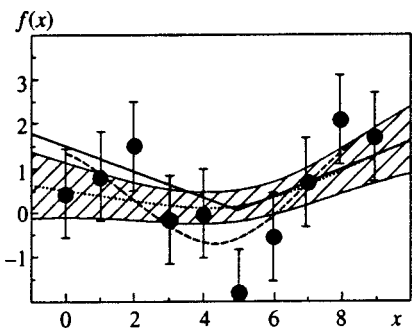

FIG. 4. Interpolation of points on a straight line (large errors, $\sigma_i = 1$). The solid curve corresponds to the ''exact'' theory, the dotted curve bordered by an error strip corresponds to the proposed approximation algorithm, and the dashed curve corresponds to interpolation by B splines.[3]

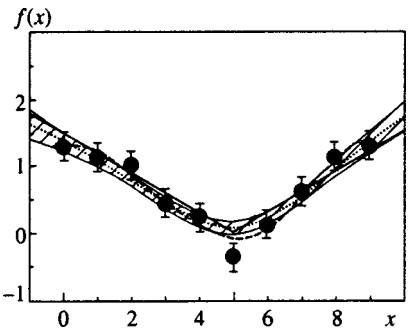

FIG. 5. Interpolation of points on a straight line (very small errors, $\sigma_i = 0.2$). The solid curve corresponds to the ''exact'' theory, the dotted curve bordered by an error strip corresponds to the proposed approximation algorithm, and the dashed curve corresponds to interpolation by B splines.[3]

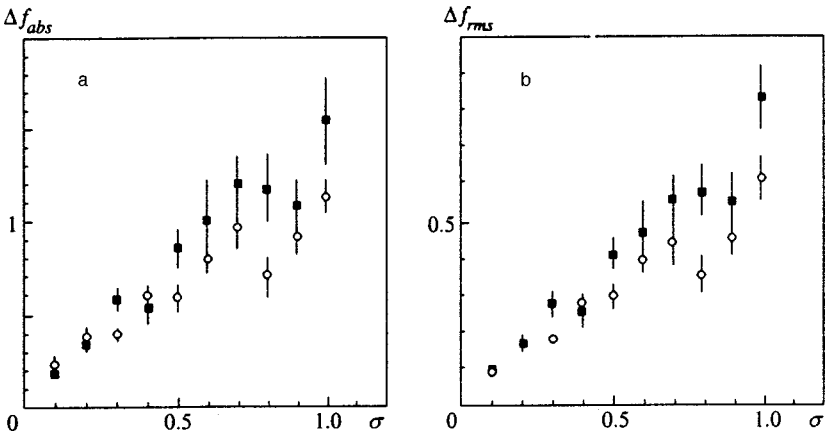

FIG. 6. Average values of the maximum $\overline{\Delta f_{\text{abs}}}$ (a) and rms $\overline{\Delta f_{\text{rms}}}$ (b) deviations of the approximating curve from the exact ''theoretical'' curve (a broken line) for two interpolation algorithms: one using B splines (■) from Ref. 3 and the algorithm proposed in this paper (○).

Smoothing algorithm	δf_{abs}	δf_{rms}
Ref. 3	0.24	0.085
This paper	0.38	0.13

The proposed algorithm falls well behind the method of interpolation by B splines both in terms of such formal estimates and from the standpoint of visual evaluation of the quality of interpolation: in the vicinity of the peak the algorithm tries to ''stretch'' the smoothing function into a straight line. This effect can be toned down by decreasing the objective value of the likelihood function: $L \to 0.7(n-2)$ [this operation corresponds to a value of the argument $q_L = L_0/(n-2) = 0.7$ in the FITAB program]. Figure 8 shows the interpolation result under the stated conditions. Here the quality of approximation in terms of absolute deviation is already higher for the proposed method than for interpolation by B splines:

Smoothing algorithm	δf_{abs}	δf_{rms}
Ref. 3	0.24	0.085
This paper	0.18	0.13

The characteristics of the approximation quality as a function of σ for this theoretical model are shown in Fig. 9. On the average the proposed algorithm exhibits better characteristics.

Even though the deficiency of the algorithm in approximating data over a large interval containing large, almost-flat segments and segments with marked irregularities is somewhat corrected by means of the argument q_L of the FITAB program, this is not an adequately effective measure. For cases involving a highly irregular function it will most likely be necessary either to generalize the smoothness tests to an integral of the square of a higher derivative or to develop a fundamentally different algorithm.

8. INFLUENCE OF ROUNDOFF ERRORS AND ESTIMATION OF THE COMPUTATION TIME

It is obvious that despite measures undertaken to allay the influence of roundoff errors in working with large matrices, the cumulative roundoff error associated with matrices having a certain number of dimensions will still severely distort the result. To investigate this effect, we have written a version of the FITAB program containing a description of all variables with a smooth point of the form real * 16. We denote the approximation obtained from this version of the program by $f_e(x)$. We then adopt the following measure of the error of the solution:

$$\Delta f_m = \max_{x_1 < x < x_n} |f(x) - f_e(x)|. \tag{46}$$

Figure 10 shows a graph of the error of the solution Δf_m as a function of the number of experimental points (which is equal to the number of dimensions of the matrices used for intermediate calculations). The last example in the preceding section (Gaussian peak against a flat background) is used as a test problem. Generally speaking, the dependence of the computational error on the number of points is not necessarily smooth, and the error depends on the specific positions of the points, on the level L_0 (in the given situation $L_0 = n-2$),

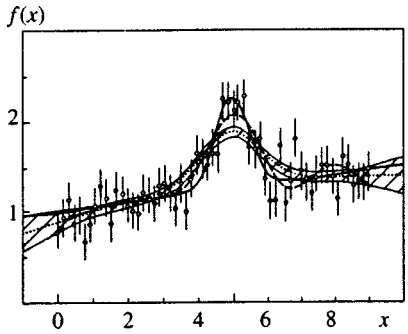

FIG. 7. Interpolation of points describing a Gaussian peak against a smooth background.

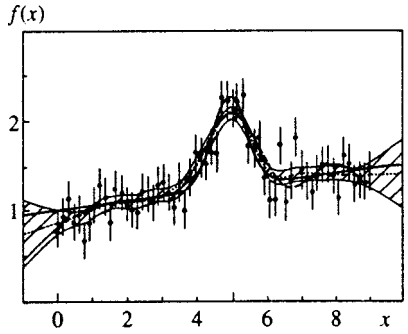

FIG. 8. Interpolation of points describing a Gaussian peak against a smooth background: $q_L = L_0/(n-2) = 0.7$.

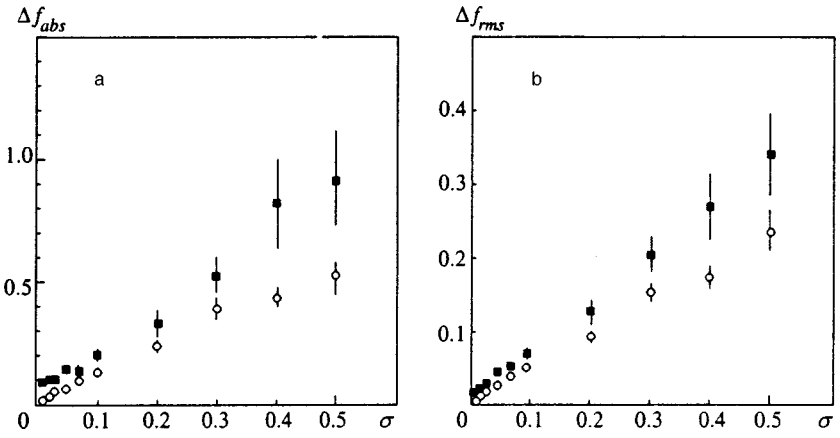

FIG. 9. Average values of the maximum $\overline{\Delta f_{\text{abs}}}$ (a) and rms $\overline{\Delta f_{\text{rms}}}$ (b) deviations of the approximating curve from the exact ''theoretical'' curve (flat background + Gaussian peak) for two interpolation algorithms: one using B splines (■) from Ref. 3 and the algorithm proposed in this paper (○).

and on the computer model (or, more precisely, on the form in which the numbers and roundoff rules are represented). The computational error has been investigated on an Alpha Server 4100 (with a clock frequency of 400 MHz) running under the Digital UNIX v4.0B (Rev. 564) operating system with a Digital FORTRAN 77 compiler. It is evident that the accumulation of roundoff errors is not very pronounced, in contrast with the simpler parametrization of the spline with respect to the coefficients at the nodes, where roundoff errors have made it impossible to run calculations with matrices of no higher than about 10–20 dimensions.

In regard to the computing time (Fig. 11), of course, the dependence on the number of points is smooth. However, not all the matrix operations in the algorithm are executed as direct cycles. The iterative method of successive rotations is used to reduce symmetric matrices to diagonal form. This operation greatly complicates the dependence of the computing time on the number of points n. If the curve in Fig. 11 is fitted by a fifth-degree polynomial, we obtain

$$t_{\text{CPU}}[\text{ms}] = \frac{n-3.2}{21} + \left(\frac{n}{45.8}\right)^2 + \left(\frac{n}{9.37}\right)^3 + \left(\frac{n}{10.5}\right)^4 + \left(\frac{n}{22.86}\right)^5. \quad (47)$$

The indicated Alpha Server 4100 has a sufficiently high speed. The same FITAB program runs 13 times slower on the Silicon Graphics Challenge-L server (clock frequency 150 MHz, IRIX 6.5 operating system, and MIPSpro v.7.2.1 compiler), 30 times slower on the VAX Station 3600 (OpenVMS v5.5 operating system and VAX FORTRAN v5.5-98 compiler), and 42 times slower on an IBM PC 486 DX/4 (clock frequency 100 MHz and Microsoft FORTRAN Visual Workbench version 1.00 compiler under the Windows 95 operating system). The comparison has been made for $n = 20$. The given coefficients permit Fig. 11 to be used to estimate computing times on other computer systems.

9. CONCLUSION

We have proposed an algorithm for the approximation of experimental data when sufficiently accurate statistical error estimates are assumed to be available, but a theoretical model is not. The method is based on the minimization of a modified likelihood function, which incorporates a generalized test of smoothness. The choice of a cubic polynomial spline of deficiency 1 with nodes at the coordinates of the experimental points as the interpolation function has been justified by investigations using methods of variational calculus. Internal parametrization of the spline in the form of a linear combination of B splines has been chosen as a means

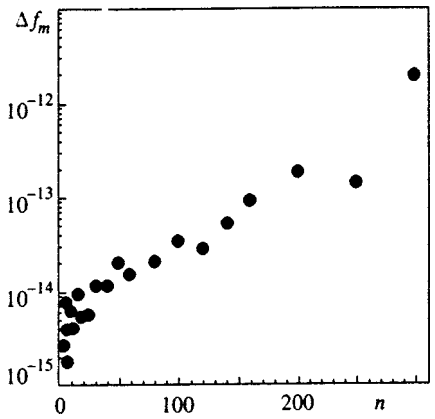

FIG. 10. Maximum deviation Δf_m of interpolation functions obtained by calculations using real * 8 and real * 16 numbers as functions of the number of points n.

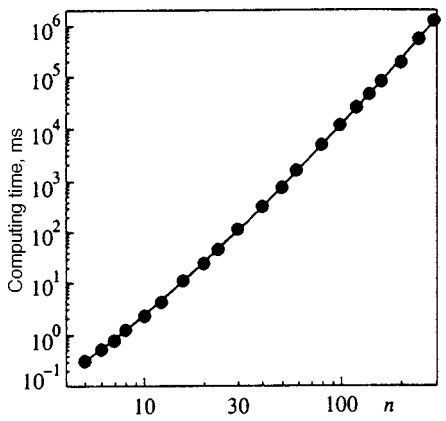

FIG. 11. Computing time of the FITAB program as a function of the number of points n (Alpha Server 4100).

of significantly diminishing the influence of roundoff errors in operations involving large-dimensional matrices.

We have presented several practical examples of interpolation with simultaneous comparisons of the quality of interpolation by means of an algorithm proposed in Ref. 3. Major differences should not be encountered here, because cubic splines are used in both cases. However, a significant difference between the two algorithms is found in the number and positions of the spline nodes.

The proposed algorithm has the drawback that the interpolation curve tends to ''stretch out into a straight line'' to the extent allowed by statistical errors of the experimental data (a similar effect is encountered in all smoothing algorithms). The problem here is attributable to the formulation of the smoothness test as an integral of the square of the second derivative of the interpolation function. If reduction of this effect is desired, it will be necessary to generalize the algorithm with a smoothness test in the form of an integral of the square of a higher derivative, which, of course, will lead to splines of higher than third degree. In this case the interpolation curve will be drawn to polynomials of higher than first degree (e.g., a parabola), to the extent allowed by experimental errors.

An interpolation program in FORTRAN 77 has been written in accordance with the proposed algorithm. The influence of roundoff errors on the computational accuracy has been investigated, and the computing time has been measured for this implementation of the algorithm has been measured over a wide range of numbers $n=5-300$.

*)E-mail: bukin@inp.nsk.su

[1] D. J. Hudson, *Lectures on Elementary Statistics and Probability*, CERN Report 63-29, CERN, Geneva (1963).
[2] L. Janossy, *Theory and Practice of the Evaluation of Measurements*, Clarendon Press, Oxford (1965).
[3] V. B. Anikeev, A. V. Popov, and V. P. Zhigunov, Nucl. Instrum. Methods Phys. Res. A **372**, 482 (1996).
[4] S. B. Stechkin and Yu. N. Subbotin, *Splines in Computational Mathematics* [in Russian], Nauka, Moscow (1976).
[5] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Prentice-Hall, Englewood Cliffs, N.J. (1963).
[6] A. D. Bukin and N. A. Grozina, Comput. Phys. Commun. **78**, 287 (1994).
[7] A. D. Bukin, IYaF Preprint No. 98-95 [in Russian] Institute of Nuclear Physics, Novosibirsk (1998).
[8] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, 2nd ed., McGraw-Hill, New York (1967).
[9] A. G. Kurosh, *Lectures in General Algebra*, Pergamon, Oxford (1965) [Russian orig., Fizmatgiz, Moscow (1963)]

Translated by James S. Wood

Teleportation of a continuous variable

S. N. Molotkov*) and S. S. Nazin

Institute of Solid-State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

(Submitted 1 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 777–792 (September 1999)

Measurements used in quantum teleportation are examined from the standpoint of the general theory of quantum-mechanical measurements. It is shown that in order to find a teleported state, it is sufficient to know only the resolution of the identity operator (positive operator-valued measure) generated by the respective instrument (the quantum operation determining the change in the state of the system as a result of the measurement) in the state space of the system, rather than the instrument itself. A protocol for quantum teleportation of the state of a system with a nondegenerate continuous spectrum based on a measurement which corresponds to a certain nonorthogonal resolution of the identity operator is proposed. © *1999 American Institute of Physics.* [S1063-7761(99)00309-1]

1. INTRODUCTION

One of the main results of quantum information theory is the possibility of teleportating an unknown quantum state by means of a classical and a distributed quantum communication channel, whose roles are played by a specially selected nonlocal (entangled state), for example, an Einstein–Podolsky–Rosen (EPR) pair.[1] The quantum teleportation of an unknown state from user A to distant user B is accomplished in the following manner.[1] User A has a state ρ_1 of quantum system 1, which is unknown to him and is to be teleported to user B (for example, a particle of spin 1/2; a generalization to the case of an arbitrary quantum system with a finite number of levels, i.e., with a finite-dimensional state space, was also given in Ref. 1). In addition, there are two other particles also of spin 1/2 (systems 2 and 3) in the spin-entangled EPR state ρ_{23}, which is such that user A has access to particle 2 and user B has access to particle 3. User A performs a certain joint measurement m_{12} on system 1 in the state ρ_1, which is unknown to him, and particle 2 from the EPR pair. As a result of the measurement, the complete system consisting of particles 1, 2, and 3 passes from the state $\rho_1 \otimes \rho_{23}$ to a new state ρ'_{123}, which depends on the result z of the measurement performed. It turns out that there are measurements m_{12}, as a result of which the state ρ'_3 of particle 3 from the EPR pair at user B (which is obtained by taking the trace of the state space of particles 1 and 2, i.e., $\rho'_3 = \mathrm{Tr}_{1,2}\rho'_{123}$) is related to the input state ρ_1 of particle 1 by a certain unitary transformation, which does not depend on ρ_1 and is determined only by the result z of the measurement m_{12} performed:

$$\rho'_3 = U_z \rho_1 \tag{1}$$

(here and below, we assume that the isomorphous state spaces of particles 1 and 3 are identical). A classical communication channel is needed for user A to communicate the measurement result z to user B. This result tells B which unitary transformation U_z^{-1} he must perform on the state ρ'_3 of particle 3 so that its state would coincide with ρ_1. We note that user A does not obtain any information on the teleported state.

In the quantum teleportation algorithm presented it is essential to utilize the fact that after performance of the measurement, the system as a whole (all three particles) is in a fully defined state ρ'_{123}, which is determined by the measurement result; the algorithm in Ref. 1 uses a so-called Bell measurement, which corresponds to a certain self-conjugate operator with a nondegenerate spectrum in four-dimensional space, and the state ρ'_{123} is easily written out in an explicit form.

The first algorithm for teleporting a continuous quantum variable (i.e., the wave function of a one-dimensional nonrelativistic spinless particle, whose state space is infinite-dimensional) was described in Ref. 2. A realistic algorithm for teleporting a single-mode electromagnetic field was subsequently proposed on the basis of that approach.[3] It was essentially assumed that in the case of an observable with a continuous spectrum, the system passes after a measurement into the state described by the ''eigenvector'' belonging to the ''eigenvalue'' of the corresponding self-conjugate operator obtained as a result of that measurement.

However, the properly posed question of the state into which a system passes after a measurement is far more complicated for a continuous variable than in the case of a discrete spectrum (see, for example, Ref. 4). The problem here is not just that in the case of a continuous spectrum there are no properly defined eigenvectors in the Hilbert state space of the system. Let us consider, for example, a certain self-conjugate operator A with a continuous spectrum Λ. Let the point z belong to this spectrum and let the system be in a certain state ρ before the measurement. How intelligent then is the question of which state ρ_z is the system in after a measurement which gives the result $r=z$? The problem stems from the fact that, according to the statistical interpretation of quantum mechanics, the very concept of a ''state'' can be applied only to a certain ensemble of identical sys-

tems and not to a single system. In the case under consideration this would appear to mean that the subensemble of systems selected after the measurement by the condition $r=z$ must be considered. However, in the case of a continuous spectrum the probability of obtaining any specific value of z is equal to zero, since a point has a measure of zero. Thus, it is simply impossible to single out a subensemble of systems which give the result $r=z$, since the probability of obtaining coinciding results in any two measurements equals zero. Therefore, the question of the meaning of ρ_z is not entirely trivial. In order to answer it, we need some facts from the general theory of quantum-mechanical measurements (see, for example, Refs. 4–6). The basic postulates and some results of that theory are presented in Sec. 2. In Sec. 3 this general theory is used to study the special class of measurements which are of interest for quantum teleportation. In Sec. 4 the protocol described in Ref. 2 for teleporting a continuous variable is examined from the standpoint of the results obtained in the preceding section. A protocol for teleporting states of a model system with a continuous spectrum, which utilizes a measurement that corresponds to a certain nonorthogonal resolution of the identity operator, is proposed in Sec. 5. Finally, the main results obtained in this work are briefly described in the last section.

2. QUANTUM-MECHANICAL MEASUREMENTS

For a quantum-mechanical state with a finite-dimensional state space $\mathscr{H}$ (in which the spectrum of any operator is purely discrete) a canonical (von Neumann) measurement of a certain observable, which corresponds to the self-conjugate operator A with the eigenvalues λ_i, where $i=1...n$, causes the system, which is initially described by the density matrix ρ, to pass into the state ρ_j (the von Neumann–Lüders reduction postulate[7,8]),

$$\rho\to\rho_j=\frac{E_j\rho E_j}{\mathrm{Tr}\{E_j\rho\}}, \tag{2}$$

if the measurement gives the result λ_j. Here E_j is the orthogonal projector onto the space corresponding to the eigenvalue λ_j, so that we have the resolution of the identity operator

$$\sum_j E_j=I, \tag{3}$$

where I is the identity operator in $\mathscr{H}$, and the spectral representation of the operator A is

$$A=\sum_j \lambda_j E_j. \tag{4}$$

The probability of obtaining the jth result equals

$$\mathrm{Prob}(\lambda_j)=\mathrm{Tr}\{\rho E_j\}=\mathrm{Tr}\{E_j\rho E_j\}. \tag{5}$$

Let us now consider the most general situation, where the set of all possible measurement results forms a certain measurable space $\mathscr{Z}$ with a measure, which will be denoted below by dz, and the quantum system S is described by the Hilbert space $\mathscr{H}$ (which is infinite-dimensional in the general case), i.e., its states obey a one-to-one correspondence to the set $K(\mathscr{H})$ of all the positive operators in $\mathscr{H}$ with a trace equal to 1 (i.e., density matrices; the operator A in $\mathscr{H}$ is called positive, if $\langle v|A|v\rangle\geqslant 0$ for any $v\in\mathscr{H}$). The set $K(\mathscr{H})$ is a subset of the space $B_1(\mathscr{H})$ of all operators with a finite trace in $\mathscr{H}$. In this case the mathematical object which completely characterizes any specific measuring procedure with the set of results $\mathscr{Z}$, to which the system S can be subjected, is the instrument[4] (or operation, according to the terminology in Ref. 5) $\mathbf{T}$. The latter is equivalent to the mapping $\Delta\to\mathbf{T}(\Delta)$ of the set Γ of all the subsets $\Delta\subset\mathscr{Z}$ which are measurable with respect to the measure dz in the set of non-trace-increasing, completely positive operators $P(B_1(\mathscr{H}))$ which map $B_1(\mathscr{H})$ into itself and satisfy the following two conditions:

1) $\mathbf{T}(\Delta)=\Sigma_j\mathbf{T}(\Delta_j)$, if $\Delta=\cup_j\Delta_j$ and $\Delta_j\cap\Delta_i=\varnothing$ for $i\neq j$ (additivity);

2) $\mathrm{Tr}\{\mathbf{T}(\mathscr{Z})\rho\}=\mathrm{Tr}\rho$ for any $\rho\in B_1(\mathscr{H})$ (normalization).

We recall that the linear mapping $\mathbf{F}$ from $B_1(\mathscr{H})$ into itself is called completely positive, if $\mathbf{F}(L)>0$ for any $L>0$ from $B_1(\mathscr{H})$, i.e., it converts positive operators from $B_1(\mathscr{H})$ into positive operators and has the additional property that if $\mathscr{H}_0$ is another Hilbert space, the mapping

$$\mathbf{F}\otimes\mathbf{I}: B_1(\mathscr{H}\otimes\mathscr{H}_0)\to B_1(\mathscr{H}\otimes\mathscr{H}_0),$$

which is assigned in elements of the type $W\otimes W_0\in B_1(\mathscr{H}\otimes\mathscr{H}_0)$ by a formula of the type

$$\mathbf{F}\otimes\mathbf{I}(W\otimes W_0)=\mathbf{F}(W)\otimes W_0$$

and is continued by linearity into the entire space $B_1(\mathscr{H}\otimes\mathscr{H}_0)$, where $\mathbf{I}$ is the identity operator in $B_1(\mathscr{H}_0)$, is also positive for any $\mathscr{H}_0$. The meaning of the instrument $\mathbf{T}$ is that for any measurable subset $\Delta\subset\mathscr{Z}$, $\Delta\in\Gamma$, the state ρ_Δ of the subensemble of systems which are initially placed in the state $\rho\in K(\mathscr{H})$ and are mapped during numerous repetitions of the measuring procedure by the condition that the measurement result $r=z$ be in Δ is (for brevity we write $\mathbf{T}(\Delta)\rho$ instead of $[\mathbf{T}(\Delta)](\rho)$)

$$\rho_\Delta=\frac{\tilde\rho(\Delta)}{\mathrm{Tr}\{\tilde\rho(\Delta)\}}=\frac{\mathbf{T}(\Delta)\rho}{\mathrm{Tr}\{\mathbf{T}(\Delta)\rho\}}\in K(\mathscr{H}),\quad \tilde\rho(\Delta)=\mathbf{T}(\Delta)\rho, \tag{6}$$

and the probability of obtaining the result $r=z\in\Delta$ following the performance of a measurement is

$$\mathrm{Prob}(z\in\Delta)=\mathrm{Tr}\{\mathbf{T}(\Delta)\rho\}=\mathrm{Tr}\{\tilde\rho(\Delta)\}. \tag{7}$$

Here and below, we use a tilde to denote ''unnormalized density matrices'' (positive operators with a trace $\leqslant 1$), which are obtained after applying the operator corresponding to the instrument under consideration $\mathbf{T}(\Delta)$ to the original density matrix ρ. In cases where no misunderstanding can arise, we shall henceforth apply the term ''density matrix'' to these operators for the sake of brevity.

It is easy to verify that for a fixed $\mathbf{T}$ formula (7) generates an affine mapping of the convex set $K(\mathscr{H})$ of the states ρ of the system S into the set of probabilistic measures ν_{Prob} in $\mathscr{Z}$: each state $\rho\in K(\mathscr{H})$ is mapped to a measure μ_ρ in $\mathscr{Z}$ so that for each set $\Delta\in\Gamma$ its measure $\mu_\rho(\Delta)$ is exactly $\mathrm{Prob}(z\in\Delta)$. As we know,[6] the set of all such mappings

$\rho \to \mu_\rho$ from $K(\mathscr{H})$ into ν_{Prob} obeys a one-to-one correspondence with the family of Hermitian operators $M(\Delta)$, $\Delta \in \Gamma$, acting in the Hilbert space $\mathscr{H}$ that satisfy the following properties:

$1'$) $M(\varnothing)=0$, $M(\mathscr{Z})=I$ (normalization);

$2'$) $M(\Delta)\geqslant 0$ (positiveness);

$3'$) $M(\Delta)=\Sigma_j M(\Delta_j)$, if $\Delta=\cup_j \Delta_j$ and $\Delta_j \cap \Delta_i=\varnothing$ for $i\neq j$ (additivity); i.e., with resolutions of the identity operator in $\mathscr{Z}$ with values in the set of positive operators in $\mathscr{H}$. In this case the measure μ_ρ of the set Δ is defined by the expression

$$\mu_\rho(\Delta)=\text{Prob}(z\in\Delta)=\text{Tr}\{\rho M(\Delta)\}. \tag{8}$$

In other words, $M(\Delta)$ defines a positive operator-valued measure. A special case of such measures is created by spectral orthogonal resolutions of the identity operator corresponding to families of spectral projectors of self-conjugate operators in $\mathscr{H}$ for which the following equality holds:

$$M(\Delta_1)M(\Delta_2)=0, \text{ if } \Delta_1\cap\Delta_2=0.$$

Measurements described by such a resolution of the identity operator are naturally called orthogonal.

Thus, if we are interested only in the probability of obtaining a particular result and leave aside the far more complicated question of the state in which the system is after a measurement, then instead of the family of operators $\mathbf{T}(\Delta) \in P(B_1(\mathscr{H}))$ it is sufficient to confine ourselves to a treatment of the positive resolutions of the identity operator $M(\Delta)\in B(\mathscr{H})$, which are related to one another in such a manner that the probability of obtaining a result $z\in\Delta$ upon the performance of a measurement for any input state ρ of the system S originally specified by formula (7) can be calculated using the operator $M(\Delta)$ from formula (8). Comparing formulas (7) and (8) to one another, we can easily see that they are consistent if and only if

$$M(\Delta)=[\mathbf{T}(\Delta)]^* I, \tag{9}$$

where the asterisk denotes a conjugate mapping from the space $B(\mathscr{H})$ into itself, and $I\in B(\mathscr{H})$ is the identity operator in $\mathscr{H}$ [we recall that the linear space of all the constrained operators $B(\mathscr{H})$ in $\mathscr{H}$ is isomorphous to the space conjugate to $B_1(\mathscr{H})$. The corresponding isomorphism is generated by the bilinear mapping

$$B(\mathscr{H})\times B_1(\mathscr{H})\to \mathbf{C}: a\in B(\mathscr{H}), b\in B_1(\mathscr{H})\to \text{Tr}\{a\cdot b\}\in\mathbf{C},$$

where $\mathbf{C}$ is the set of complex numbers].

In the case of a canonical measurement of the observable A (i.e., a discrete spectrum) described by formulas (2)–(4) in the finite-dimensional space $\mathscr{H}$, the space $\mathscr{Z}$ coincides with the finite set of eigenvalues λ_i, $i=1...n$, of the operator A, the set Γ consists of all the subsets of $\mathscr{Z}$, and the operators $\mathbf{T}(\{\lambda_j\})$ and $M(\{\lambda_j\})$ for the single-point sets $\{\lambda_j\}$ have the form

$$\mathbf{T}(\{\lambda_j\})\rho=E_j\rho E_j, \quad M(\{\lambda_j\})=E_j. \tag{10}$$

The family of operators $\mathbf{T}(\Delta)$ clearly gives a far more complete description of the measurement process than does the corresponding resolution of the identity operator $M(\Delta)$, since the former not only permits calculation of the statistics of measurement results, but also specifies the state of the system after a measurement (6). In the general case, the same resolution of the identity operator $M(\Delta)$ can be generated by different instruments $\mathbf{T}_1\neq\mathbf{T}_2$.

It then turns out[4] that in the case of $\mathscr{Z}=\mathbf{R}$ (the real straight line), for any fixed ρ we have the following integral representation for $\tilde{\rho}(\Delta)=\mathbf{T}(\Delta)\rho$:

$$\tilde{\rho}(\Delta)=\mathbf{T}(\Delta)\rho=\int_\Delta \rho_z \text{Tr}\{\rho M(dz)\}, \tag{11}$$

where ρ_z is a function of the space of measurement results $\mathscr{Z}$ in the space of density matrices $K(\mathscr{H})$, and $\text{Tr}\{\rho M(dz)\}$ is the ''density'' of the measure μ_ρ (8) in $\mathscr{Z}$, i.e.,

$$\mu_\rho(\Delta)=\text{Prob}(z\in\Delta)=\text{Tr}\{\rho M(\Delta)\}=\int_\Delta \text{Tr}\{\rho M(dz)\}, \tag{12}$$

$$\mu_\rho(\Delta)=\int_\Delta d\mu_\rho(z), \quad d\mu_\rho(z)=\text{Tr}\{\rho M(dz)\}. \tag{13}$$

The function ρ_z thus defined can already be interpreted as ''the state of the system after a measurement which gave the result z.'' This does not contradict the statistical interpretation of quantum mechanics, since, in reality, ρ_z only serves as a convenient auxiliary instrument, which permits calculation of the final state of the system after a measurement. The physical interpretation of formula (11) is perfectly clear, since $\text{Tr}\{\rho M(dz)\}$ is the probability of obtaining a result in the vicinity dz of the point z in a measurement. A representation of the type (11) is important for us, because in the case of teleportation the state of the system after a measurement is corrected using a certain unitary transformation U_z, which depends on the result obtained z. In this case the subensemble of systems mapped by the condition $z\in\Delta$ after the unitary correction can clearly be described by the density matrix

$$\tilde{\rho}_{U,\Delta}=\int_\Delta U_z\rho_z U_z^+ \text{Tr}\{\rho M(dz)\}; \tag{14}$$

therefore, the introduction of the function ρ_z is a natural step in an attempt to extend the algorithm for teleporting a state of a finite-dimensional quantum system described in Ref. 1 to the case of a continuous variable.

3. MEASUREMENTS USED IN QUANTUM TELEPORTATION

Let us now consider the measurements which are used in quantum teleportation algorithms from the standpoint of the general quantum-mechanical theory presented above. Let a measurement corresponding to the instrument $\mathbf{T}_{12}$ be performed on particles 1 and 2. Then, with respect to the entire system, including particle 3, this measurement corresponds to the instrument $\mathbf{T}_{123}(\Delta)=\mathbf{T}_{12}(\Delta)\otimes\mathbf{I}_3$, where $\mathbf{I}_3$ is the identity operator in $B_1(\mathscr{H}_3)$. It therefore follows that after a joint measurement performed on the first and second systems, the subensemble of systems mapped by the condition $z\in\Delta$, where $\Delta\subset\mathscr{Z}$ and $\Delta\in\Gamma$ (we have still not concretized the space of results $\mathscr{Z}$), is described by the density matrix

$$\rho'_{123,\Delta}=\frac{\mathbf{T}_{123}(\Delta)\rho}{\mathrm{Tr}_{1,2,3}\{\mathbf{T}_{123}(\Delta)\rho\}}, \tag{15}$$

and the probability that z falls in Δ is $\mathrm{Tr}_{1,2,3}\{\mathbf{T}_{123}(\Delta)\rho\}$. The reduced density matrix describing the state of system 3 has the form

$$\rho'_{3,\Delta}=\frac{\mathrm{Tr}_{1,2}\{\mathbf{T}_{123}(\Delta)\rho\}}{\mathrm{Tr}_{1,2,3}\{\mathbf{T}_{123}(\Delta)\rho\}}. \tag{16}$$

Here we have a special case of the following more general situation. Let there be a composite system S, which consists of two systems A and B and is in the state ρ_{AB} (in the case of teleportation the role of the system A is played by particles 1 and 2 considered together, and the role of the system B is played by particle 3). Next, let a measurement be performed on the system A using the instrument $\mathbf{T}_A$, and let us find the state $\rho'_{B,\Delta}$ of the system B after the measurement (here and in the following the prime sign indicates that the state under consideration is the state of the particular system immediately after the measurement). The instrument $\mathbf{T}_{AB}$, which describes the change in the state of the entire system, is $\mathbf{T}_A\otimes\mathbf{I}_B$, and, therefore,

$$\rho'_{B,\Delta}=\frac{\mathrm{Tr}_A\{\mathbf{T}_{AB}(\Delta)\rho_{AB}\}}{\mathrm{Tr}_{AB}\{\mathbf{T}_{AB}(\Delta)\rho_{AB}\}}. \tag{17}$$

Let us now examine the numerator in this formula, which we denote by $\tilde{\rho}'_{B,\Delta}=\mathrm{Tr}_A\{\mathbf{T}_{AB}(\Delta)\rho_{AB}\}$ in accordance with the conventions adopted above (then the probability that the measurement result z belongs to Δ is $\mathrm{Tr}_B\tilde{\rho}'_{B,\Delta}$). Let u_B be an arbitrary operator from $B(\mathscr{H})$. We calculate the trace $\mathrm{Tr}_B\{u_B\tilde{\rho}'_B\}$ (for brevity we omit the subscript Δ everywhere):

$$\begin{aligned}
\mathrm{Tr}_B\{u_B\tilde{\rho}'_B\}&=\mathrm{Tr}_B\{u_B\mathrm{Tr}_A\{\mathbf{T}_A\otimes\mathbf{I}_B\rho_{AB}\}\}\\
&=\mathrm{Tr}_B\{\mathrm{Tr}_A\{I_A\otimes u_B\cdot\mathbf{T}_A\otimes\mathbf{I}_B\rho_{AB}\}\}\\
&=\mathrm{Tr}_{AB}\{I_A\otimes u_B\cdot\mathbf{T}_A\otimes\mathbf{I}_B\rho_{AB}\}\\
&=\mathrm{Tr}_{AB}\{[(\mathbf{T}_A\otimes\mathbf{I}_B)^*I_A\otimes u_B]\cdot\rho_{AB}\}\\
&=\mathrm{Tr}_{AB}\{[(\mathbf{T}_A^*I_A)\otimes\mathbf{I}_B^*u_B]\cdot\rho_{AB}\}\\
&=\mathrm{Tr}_{AB}\{[M_A\otimes u_B]\cdot\rho_{AB}\}\\
&=\mathrm{Tr}_{AB}\{[(M_A\otimes I_B)\cdot(I_A\otimes u_B)]\cdot\rho_{AB}\}\\
&=\mathrm{Tr}_{AB}\{[(I_A\otimes u_B)\cdot(M_A\otimes I_B)]\cdot\rho_{AB}\}\\
&=\mathrm{Tr}_B\{\mathrm{Tr}_A\{[(I_A\otimes u_B)\cdot(M_A\otimes I_B)]\cdot\rho_{AB}\}\}\\
&=\mathrm{Tr}_B\{u_B\mathrm{Tr}_A\{(M_A\otimes I_B)\cdot\rho_{AB}\}\}.
\end{aligned} \tag{18}$$

Therefore,

$$\tilde{\rho}'_{B,\Delta}=\mathrm{Tr}_A\{\mathbf{T}_{AB}(\Delta)\rho_{AB}\}=\mathrm{Tr}_A\{(M_A(\Delta)\otimes I_B)\cdot\rho_{AB}\}. \tag{19}$$

Thus, in order to find the state of the system B after a measurement performed on the system A, it is sufficient to know only the resolution of the identity operator in $\mathscr{H}_A$ generated by T_A in $\mathscr{Z}$, rather than the instrument $\mathbf{T}_A$ itself.

We note that the machinery of quantum operators was probably first applied to teleportation in Ref. 9, where the simple case of ''ideal'' teleportation with a discrete space of measurement outcomes $\mathscr{Z}$ was considered. In that case the change in the state of the system caused by a measurement is described by an instrument of the type

$$\rho\to A_i\rho A_i^+, \tag{20}$$

where A_i is a certain positive operator, and $i=1,2\ldots$ labels the various measurement outcomes, i.e., points in $\mathscr{Z}$. However, the teleported state was expressed in terms of the operators A_i, which completely assign the instrument itself.

We are interested in the possibility of representing $\tilde{\rho}'_{B,\Delta}$ in the form

$$\tilde{\rho}'_{B,\Delta}=\int_\Delta\rho_{z,B}d\mu_{\rho_{AB}}(z), \tag{21}$$

where $\rho_{z,B}\in K(\mathscr{H}_B)$, and the measure $d\mu_{\rho_{AB}}(z)$ describes the probability density of the measurement result falling in the vicinity of the point z, i.e., satisfies the condition

$$\mathrm{Tr}_B\tilde{\rho}'_{B,\Delta}=\int_\Delta d\mu_{\rho_{AB}}(z). \tag{22}$$

Formally, such a representation can easily be found if the measure $\mu_{\rho_{AB}}$ is absolutely continuous relative to the input measure dz in $\mathscr{Z}$, and the matrix elements of the operator $M_A(\Delta)$ acting on a certain orthonormalized basis $|\varphi_{nA}\rangle$ of the system A can be represented in the form

$$\langle\varphi_{mA}|M_A(\Delta)|\varphi_{nA}\rangle=\int_\Delta dzF_{mn}(z), \tag{23}$$

where $F_{mn}(z)$ are certain complex-valued functions in $\mathscr{Z}$ [for example, if the measurement M corresponds to the simultaneous measurement of the complete system of commuting observables with a continuous spectrum, since in that case $\mathscr{H}_A=L^2(\mathscr{Z})$, and $\mathscr{Z}$ itself is the product of the spectra of the operators appearing in this system, so that $F_{mn}(z)=\varphi_{mA}(z)^*\psi_{nA}(z)$]. In fact, in this case

$$\begin{aligned}
\tilde{\rho}'_{B,\Delta}&=\mathrm{Tr}_A\{(M_A(\Delta)\otimes I_B)\cdot\rho_{AB}\}\\
&=\sum_{mn}\langle\varphi_{mA}|M_A(\Delta)|\varphi_{nA}\rangle\rho_{nm,B}\\
&=\sum_{mn}\int_\Delta dzF_{mn}(z)\rho_{nm,B}=\int_\Delta dz\left[\sum_{mn}F_{mn}(z)\rho_{nm,B}\right]\\
&=\int_\Delta dz\,\tilde{\rho}_{z,B},
\end{aligned} \tag{24}$$

where the operator $\rho_{nm,B}$ in $\mathscr{H}_B$ is obtained from ρ_{AB} by taking the ''partial matrix element'' with respect to the vectors φ_{nA} and φ_{mA} from $\mathscr{H}_A$,

$$\rho_{nm,B}=\langle\varphi_{nA}|\rho_{AB}|\varphi_{mA}\rangle, \tag{25}$$

and

$$\tilde{\rho}_{z,B}=\sum_{mn}F_{mn}(z)\rho_{nm,B}. \tag{26}$$

Therefore,

$$\mathrm{Tr}_B\{\tilde{\rho}'_{B,\Delta}\}=\int_\Delta d\mu_{\rho_{AB}}(z)=\int_\Delta dz\,\mathrm{Tr}_B\{\tilde{\rho}_{z,B}\}. \tag{27}$$

Thus, by multiplying and dividing the integrand in the last integral in Eq. (24) by $H(z)=\mathrm{Tr}\{\tilde{\rho}_{z,B}\}>0$ we obtain formula (21), where

$$\rho_{z,B}=\frac{\tilde{\rho}_{z,B}}{\mathrm{Tr}\{\tilde{\rho}_{z,B}\}}=\frac{\tilde{\rho}_{z,B}}{H(z)}, \tag{28}$$

so that $\mathrm{Tr}\{\rho_{z,B}\}=1$, and $d\mu\rho_{AB}(z)=H(z)dz$, i.e., $H(z)$ is the Radon–Nicodim derivative of $d\mu_{\rho_{AB}}(z)$ with respect to dz. We shall not dwell on substantiating the correctness of the rearrangement of the summation of the infinite series and integration in (24) and other such operations, since in the concrete cases considered in the remainder of this paper the integral representation of type (24) follows from the concrete form of the operators $M(\Delta)$.

4. TELEPORTATION USING AN ORTHOGONAL MEASUREMENT

As an illustration of the general scheme described above, we first consider the teleportation of an unknown quantum state $|\psi\rangle$ of a one-dimensional nonrelativistic spinless particle in the coordinate representation. In order to avoid the complications associated with consideration of the symmetry relative to interchange of the particles, we assume that all three particles are different. It is sufficient to consider the case where the input state of particle 1 is pure:

$$\rho_1=\rho_\psi=|\psi;1\rangle\langle\psi;1|,\quad |\psi;1\rangle=\int_{-\infty}^{\infty}dx\,\psi(x)|x;1\rangle. \tag{29}$$

The entangled state of particles 2 and 3 is chosen in the form of an EPR state (with an infinite norm)

$$\rho_{23}=|\psi_{23}\rangle\langle\psi_{23}|,\ \ |\psi_{23}\rangle=\int_{-\infty}^{\infty}dx|x;2\rangle|x;3\rangle, \tag{30}$$

which can be represented as the limit of the normalized state

$$|\Psi_{23}\rangle=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dxdy\Psi(x,y)|x;2\rangle|y;3\rangle, \tag{31}$$

where $\Psi(x,y)\to\delta(x-y)$ [in the momentum representation $\Psi_{23}(p_1,p_2)\to\delta(p_1+p_2)$]. Formally, the state (30) is an eigenvector of the coordinate difference operator of the second and third particles: $(X_2-X_3)|\psi_{23}\rangle=0$.

Let us consider a joint measurement of one of the particles in the EPR pair and of the system in the unknown state. Such a measurement can be defined by the following resolution of the identity operator:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E_{12}(dXdP)=I, \tag{32}$$

$$E_{12}(dXdP)=|\Phi_{XP}\rangle\left\langle\Phi_{XP}\right|\frac{dXdP}{2\pi} \tag{33}$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dx'e^{iP(x-x')}$$

$$\times|x+X;1\rangle|x;2\rangle\langle x'+X;1|\langle x';2|dXdP, \tag{34}$$

where

$$|\Phi_{XP}\rangle=\int_{-\infty}^{\infty}dxe^{iPx}|x+X;1\rangle|x;2\rangle. \tag{35}$$

We note that the state (35) is formally a common eigenvector for the pair of commuting observables X_1-X_2 and P_1+P_2 (the coordinate difference and the total momentum), which form a complete set of commuting operators in the state space of the two particles:

$$(X_1-X_2)|\Phi_{XP}\rangle=X|\Phi_{XP}\rangle,\ \ (P_1+P_2)|\Phi_{XP}\rangle=P|\Phi_{XP}\rangle.$$

Therefore, teleportation with ρ_{23} of the form (30) and the measurement (33) coincides exactly with the algorithm in Ref. 2. In the present case the space of measurement results $\mathscr{Z}$ is the set of the ordered pairs (X,P) of the points $(-\infty<X<\infty,\ -\infty<P<\infty)$, which form the $\mathbf{R}^2$ plane. The latter is the direct product of the two copies of the real straight line $\mathbf{R}_X$ and $\mathbf{R}_P$, which correspond to the ''coordinate'' X and the ''momentum'' P: $\mathscr{Z}=\mathbf{R}_X\times\mathbf{R}_P$.

The exact meaning of formula (33) is that the matrix elements of the positive operator $E(\Delta)$, as applied to the set Δ, can be calculated from the formula

$$\langle\Phi|E_{12}(\Delta)|\Psi\rangle=\int_\Delta\frac{dXdP}{2\pi}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dx'e^{iP(x-x')}\Phi^*$$

$$\times(x+X,x)\Psi(x'+X,x'), \tag{36}$$

which is analogous to formula (23).

Simple calculations show that the teleported density matrix in channel 3 takes the form

$$\tilde{\rho}'_{3,\Delta}=\mathrm{Tr}_{1,2}\{(\rho_1\otimes\rho_{23})E_{12}(\Delta)\}=\int_\Delta\rho_{XP}\frac{dXdP}{2\pi}, \tag{37}$$

where

$$\rho_{XP}=|\psi_{XP};3\rangle\langle\psi_{XP};3|,\quad \psi_{XP}(x)=e^{iPx}\psi(x+X). \tag{38}$$

Since

$$\mathrm{Tr}_3\{\rho_{XP}\}=\int_{-\infty}^{\infty}dx|\psi(x+X)|^2=1, \tag{39}$$

it is clear that the probability density of obtaining values in the vicinity of the point (X,P) in the interval (dX,dP) in a measurement is $1/2\pi$ and does not depend on $|\psi;1\rangle$, so that the measurement does not provide any information on the teleported state. The total probability of obtaining a pair (X,P) at all is infinite because of the unnormalized nature of the state (30).

It follows from formulas (37) and (38) that by applying the unitary transformation

$$U_{XP}:\ \ \psi(x)\to e^{iP(x-X)}\psi(x-X), \tag{40}$$

which depends only on the result of the measurement performed on particles 1 and 2, to particle 3, we obtain a state coinciding with the input state of particle 1 in channel 3, i.e., we accomplish the teleportation of the state of system 1. We note that in this example the unitary correction (which does not depend on ρ_1) of the state of the third particle to ρ_1 is possible for any input state ρ_1 and any measurement outcome, i.e., for any pair (X,P). However, generally speaking, it is reasonable to also consider teleportation algorithms which permit the teleportation not of all the possible states of particle 1, but only of a certain subset $K'(\mathscr{H}_1)$, for example, the states belonging to a certain subspace $\mathscr{H}_1' \subset \mathscr{H}_1$ (Ref. 9) (an example is considered in the next section). In addition, the requirement that the necessary unitary correction U_z exists for any measurement outcome is likewise not obligatory. In fact, the entire space of possible measurement outcomes $\mathscr{Z}$ can always be divided into two nonintersecting subsets $\mathscr{Z}_1$ and $\mathscr{Z}_2$, i.e.,

$$\mathscr{Z}=\mathscr{Z}_1 \cap \mathscr{Z}_2=\varnothing, \quad \mathscr{Z}=\mathscr{Z}_1 \cup \mathscr{Z}_2,$$

in the following manner: an arbitrary point $z \in \mathscr{Z}$ belongs to the set $\mathscr{Z}_1$ if and only if the unitary transformation U_z with the properties needed exists. A sufficient condition for teleportation is then a nonzero measure $\mu_\rho(\mathscr{Z}_1)$ for all $\rho \in K'(\mathscr{H}_1)$. The teleportation algorithm has the following form: an ensemble of systems representing the input state ρ_1 is subjected to a joint measurement m_{12} with particle 2. If a result $z \in \mathscr{Z}_2$ is obtained, the respective copy of system 3 is discarded. If $z \in \mathscr{Z}_1$, system 3 is subjected to the unitary correction U_z. Then the subensemble of particles 3 thus discarded and corrected is in the same input state ρ_1.

5. TELEPORTATION USING A NONORTHOGONAL MEASUREMENT

Let us now consider the example of the teleportation of an unknown state using a measurement described by a nonorthogonal resolution of the identity operator. We consider a model quantum system, whose Hamiltonian has a purely continuous degenerate spectrum coinciding with the interval $(0,+\infty)$ (one example is a free nonrelativistic one-dimensional spinless particle, whose attainable states are confined by the condition that their resolution contains plane waves traveling in one arbitrarily chosen direction). Thus, we assume that an arbitrary pure state of system 1 is assigned by a wave function defined on the positive semiaxis:

$$|\psi;1\rangle=\int_0^\infty \psi(E)|E;1\rangle dE, \quad \langle E|E'\rangle=\delta(E-E'). \quad (41)$$

The EPR state can be chosen in the energy representation in, for example, the form

$$|\psi_{23}\rangle=\int_0^{\varepsilon_0} d\varepsilon |\varepsilon;2\rangle|\varepsilon_0-\varepsilon;3\rangle. \quad (42)$$

Such an EPR pair can be regarded as the limit of the normalized state

$$|\Psi_{23}\rangle=\int_0^{\varepsilon_0}\int_0^{\varepsilon_0} d\varepsilon_1 d\varepsilon_2 \psi(\varepsilon_1,\varepsilon_2)|\varepsilon_1;1\rangle|\varepsilon_2;2\rangle, \quad (43)$$

where $\psi(\varepsilon_1,\varepsilon_2)\to\delta(\varepsilon_1+\varepsilon_2-\varepsilon_0)$. Such a state is obtained following parametric energy down-conversion, if the pump frequency equals ε_0. Formally, the EPR state can also be chosen in the form $\psi(\varepsilon_1,\varepsilon_2)\to\delta(\varepsilon_1-\varepsilon_2)$; however, it is then not very clear how such a state could be realized experimentally.

Let us now consider the joint measurement $M_{12}(d\Omega dT)$ of particles 1 and 2, which can be represented in the form of a nonorthogonal resolution of the identity operator:

$$\begin{aligned} M_{12}(d\Omega dT) &= \frac{1}{\pi}\left(\int_{-\Omega}^{\Omega} d\omega e^{i\omega T}|\Omega+\omega;1\rangle|\Omega-\omega;2\rangle\right) \\ &\times\left(\int_{-\Omega}^{\Omega} d\omega' e^{-i\omega' T}\langle\Omega+\omega';1|\langle\Omega-\omega';2|\right) d\Omega dT \end{aligned} \quad (44)$$

$$\begin{aligned} &= \frac{1}{\pi}\int_{-\Omega}^{\Omega}\int_{-\Omega}^{\Omega} d\omega d\omega' e^{i(\omega-\omega')T}|\Omega+\omega;1\rangle \\ &\times|\Omega-\omega;2\rangle\langle\Omega+\omega';1|\langle\Omega-\omega';2|d\Omega dT. \end{aligned} \quad (45)$$

Here Ω and T vary in the intervals $\mathbf{R}_\Omega^+=(0;+\infty)$ and $\mathbf{R}_T=(-\infty;+\infty)$, respectively, so that the space of all possible measurement results is $\mathscr{Z}=\mathbf{R}_\Omega^+\times\mathbf{R}_T$. The quantities Ω and ω have the meaning of the half-sum and the half-difference of the energies (we do not distinguish between frequency and energy) of two particles, for example, the photons in a biphoton. Such a measurement, which, in a certain sense, is intermediate between measures of the frequency and the time parameter for two-particle states, can, in principle, be performed experimentally for two-particle states using parametric energy up-conversion.[11]

It is not difficult to show that $M_{12}(d\Omega dT)$ is, in fact, a resolution of the identity operator:

$$\begin{aligned} &\int_0^\infty\int_{-\infty}^{\infty} M_{12}(d\Omega dT) \\ &= \frac{1}{\pi}\int_0^\infty d\Omega\int_{-\infty}^{\infty} dT\int_{-\Omega}^{\Omega} d\omega\int_{-\Omega}^{\Omega} d\omega' e^{i(\omega-\omega')T} \\ &\times|\Omega+\omega;1\rangle|\Omega-\omega;2\rangle\langle\Omega+\omega';1|\langle\Omega-\omega';2| \\ &= 2\int_0^\infty d\Omega\int_{-\Omega}^{\Omega} d\omega\int_{-\Omega}^{\Omega} d\omega'\,\delta(\omega-\omega') \\ &\times|\Omega+\omega;1\rangle|\Omega-\omega;2\rangle\langle\Omega+\omega';1|\langle\Omega-\omega';2| \\ &= \int_0^\infty d\omega_1\int_0^\infty d\omega_2|\omega_1;1\rangle|\omega_2;2\rangle\langle\omega_1;1|\langle\omega_2;2| = I_{12}, \end{aligned}$$

where $\omega_1=\Omega+\omega$ and $\omega_2=\Omega-\omega$.

It can easily be shown that the teleported density matrix now has the form

$$\tilde{\rho}'_{3,\Delta}=\mathrm{Tr}_{1,2}\{(\rho_1\otimes\rho_{23})M_{12}(\Delta)\}=\int_\Delta \rho_{\Omega T}\frac{d\Omega dT}{\pi},$$

$$\rho_{\Omega T}=|\psi_{\Omega T};3\rangle\langle\psi_{\Omega T};3|, \quad (46)$$

where (for brevity we write $|\psi_3\rangle$ instead of $|\psi_{\Omega T};3\rangle$):

$$|\psi_{\Omega T};3\rangle=|\psi_3\rangle=\int_{\varepsilon_0-\min\{\varepsilon_0,2\Omega\}}^{\varepsilon_0} d\varepsilon e^{-i(2\Omega-\varepsilon_0+\varepsilon)T} \times\psi(2\Omega-\varepsilon_0+\varepsilon)|\varepsilon;3\rangle. \quad (47)$$

The probability of obtaining values in the interval $(\Omega,\Omega+d\Omega;T,T+dT)$ in measurements equals

$$\mathrm{Tr}\{\tilde{\rho}'_{d\Omega dT}\}=\mathrm{Tr}_{1,2,3}\{(\rho_1\otimes\rho_{23})M_{12}(d\Omega dT)\} =\frac{d\Omega dT}{\pi}\int_{\varepsilon_0-\min\{\varepsilon_0,2\Omega\}}^{\varepsilon_0}|\psi(2\Omega-\varepsilon_0+\varepsilon)|^2 d\varepsilon. \quad (48)$$

We note that the corresponding probability density does not depend on T. Since T varies in an infinite range, the total probability turns out to be infinite, as in the preceding section. Formally, this is due to the fact that the state (42) has an infinite norm. However, this does not lead to any difficulties, since knowledge of the relative probabilities of the occurrence of the various events is fully sufficient for obtaining physically intelligible results.

We now assume that we know that the carrier of the function ψ of system 1 is concentrated in a certain interval $[E_{\min},E_{\max}]$, i.e., $\psi(E)=0$ at $E>E_{\max}$ and $E<E_{\min}$. The probability density (48) then begins to depend on Ω. For example, it vanishes at $2\Omega>E_{\max}+\varepsilon_0$, since in this case ψ is identically equal to zero over the entire integration interval. The appearance of the carrier of ψ in the integration interval in (47) is clearly a condition for the realization of exact teleportation. In this case the probability density for obtaining a particular value of Ω does not depend on $|\psi;1\rangle$, since the integral in (48) is identically equal to unity by virtue of the normalization of $|\psi;1\rangle$.

It is convenient to perform the further analysis separately for the cases of $\varepsilon_0>E_{\max}$ and $\varepsilon_0<E_{\max}$. Let us first consider the case of $\varepsilon_0>E_{\max}$. If a measurement gives a result $2\Omega<\varepsilon_0$ (case 1a), the state of system 3 will have the form $|\psi_3\rangle\langle\psi_3|$, where

$$|\psi_3\rangle=\int_{\gamma}^{\varepsilon_0} d\varepsilon e^{-i(\varepsilon-\gamma)T}\psi(\varepsilon-\gamma)|\varepsilon;3\rangle,\quad \gamma=\varepsilon_0-2\Omega. \quad (49)$$

The argument of ψ in the integrand varies in the range from 0 to 2Ω. Hence it follows that teleportation of the state ψ is possible only if its carrier $[E_{\min},E_{\max}]\subset[0,2\Omega]$, i.e., if $E_{\max}<2\Omega$. Thus, $[E_{\max},\varepsilon_0]\subset\mathscr{Z}_1$ (we omit the trivial direct multiplier $\mathbf{R}_T$ in $\mathscr{Z}_1$, since nothing depends on the value of T).

If a measurement gives a result $2\Omega>\varepsilon_0$ (case 1b), the state of system 3 will have the form $|\psi_3\rangle\langle\psi_3|$, where

$$|\psi_3\rangle=\int_{0}^{\varepsilon_0} d\varepsilon e^{-i(\varepsilon+\gamma)T}\psi(\varepsilon+\gamma)|\varepsilon;3\rangle,\quad \gamma=2\Omega-\varepsilon_0. \quad (50)$$

Now the argument of ψ in the integrand varies in the range from γ to 2Ω, and teleportation of the state ψ is possible only if its carrier $[E_{\min},E_{\max}]\subset[\gamma,2\Omega]$, i.e., if $\gamma<E_{\min}$, or, stated differently, $2\Omega<\varepsilon_0+E_{\min}$ (the condition $E_{\max}<2\Omega$ is satisfied automatically, since $2\Omega>\varepsilon_0>E_{\max}$). Thus, $[\varepsilon_0,\varepsilon_0+E_{\min}]\subset\mathscr{Z}_1$. Combining cases 1a and 1b, we obtain $\mathscr{Z}_1=[E_{\max},\varepsilon_0+E_{\min}]$.

It is seen from Eqs. (49) and (50) that in cases 1a and 1b system 3 goes over to a state which is identical to the state of system 1 before the measurement, if, immediately after the measurement, it is subjected to the unitary transformations

$$\psi(\varepsilon)\to\tilde{\psi}(\varepsilon)=\begin{cases}\psi(\varepsilon), & \text{if } \varepsilon>\varepsilon_0,\\ \psi(\varepsilon+\gamma)e^{i\varepsilon T}, & \text{if } 0<\varepsilon<2\Omega,\\ \psi(\varepsilon-2\Omega), & \text{if } 2\Omega<\varepsilon<\varepsilon_0,\end{cases} \quad (51)$$

and

$$\psi(\varepsilon)\to\tilde{\psi}(\varepsilon)=\begin{cases}\psi(\varepsilon), & \text{if } \varepsilon>2\Omega,\\ \psi(\varepsilon+\gamma)e^{i\varepsilon T}, & \text{if } \gamma<\varepsilon<2\Omega,\\ \psi(\varepsilon-2\Omega), & \text{if } 0<\varepsilon<\gamma,\end{cases} \quad (52)$$

respectively. Let us now consider the situation in which $\varepsilon_0<E_{\max}$. In this case measurements which gave a result $2\Omega<\varepsilon_0$, are sure to be unsuitable for teleportation, since the range of variation of the argument of ψ in (49) does not cover the carrier of ψ. However, if a measurement gives a result $2\Omega>\varepsilon_0$, then, just as in case 1b, teleportation is possible [using the unitary transformation (50)], if $[E_{\min},E_{\max}]\subset[\gamma,2\Omega]$, i.e., if the conditions $\gamma<E_{\min}$ (i.e., $2\Omega<\varepsilon_0+E_{\min}$) and $E_{\max}<2\Omega$ are satisfied simultaneously (now this inequality imposes another additional condition, and it is not satisfied automatically). In order that a certain range of values of Ω, in which the conditions $2\Omega<\varepsilon_0+E_{\min}$ and $E_{\max}<2\Omega$ are satisfied simultaneously, would exist, the inequality $E_{\max}<E_{\min}+\varepsilon_0$ or, stated differently, $\varepsilon_0>E_{\max}-E_{\min}$ must be satisfied. Once again, $\mathscr{Z}_1=[E_{\max},\varepsilon_0+E_{\min}]$. Thus, in the proposed scheme teleportation is possible if and only if the width of the spectrum of the EPR pair (42) exceeds the spectral width of the carrier of ψ.

We note that the question of the teleportation of a broad-band one-photon wave packet was first considered in Refs. 10 and 11. In addition, the algorithm for teleporting a single-mode electromagnetic field using a squeezed state[3] was recently generalized to the case of a broad-band input state,[12] whose spectral density was assumed to be concentrated in the vicinity of the half-frequency of the pump field generating the squeezed state. Unlike the algorithm described above, the scheme in Ref. 12 is based on orthogonal measurements. From the physical standpoint the nonorthogonal measurement (44) naturally appears when the states of the system are considered in the energy representation: just as in the originally proposed teleportation scheme, which was described in the coordinate representation,[2] a simultaneous measurement of the coordinate and the momentum is employed. It is natural to presume that a similar procedure can be implemented using a measurement of the energy and its conjugate, i.e., the time. However, because no self-conjugate operator corresponds to the observed time in quantum mechanics, the measurement obtained is nonorthogonal (an EPR pair in which the states of the particles are energy-entangled, rather than coordinate-entangled, is, of course, used in this case).

We note that the teleportation of a quantum state described by dynamic variables (x,p) was investigated in Ref. 3 (the unknown state in Ref. 3 corresponds to a single-mode photon state) for the case of a nonideal EPR pair (a squeezed state). The nonideality of the EPR correlations leads to a decrease in teleportation fidelity. As can be seen from the example based on an orthogonal measurement, unconditional exact teleportation (fidelity$=1$) can be achieved with a singular EPR state. Unconditional teleportation refers here to a situation in which any measurement outcome leads to exact teleportation. In the case of a nonorthogonal measurement considered above, unconditional exact teleportation is impossible even for a singular EPR pair: for some measurement outcomes there is no unitary transformation whose application to the teleported state would transform it into an exact copy of the input state; such outcomes must be discarded. Exact teleportation occurs for the remaining measurement outcomes.

In teleportation experiments there can be a situation where instead of a measurement which theoretically leads to exact and unconditional teleportation for any outcome, a certain approximation of it is actually realized, and the teleportation becomes conditional (even if it is assumed that an ideal EPR pair is used in the experiments). Formally, any measurement is described by a resolution of the identity operator; when any particular resolution of the identity operator is realized experimentally, an interaction with the measuring device must be chosen which is such that the measurement outcomes would produce the probability distribution specified by the particular resolution of the identity operator and all the possible measurement results would be exhausted by these outcomes. As a rule, it is difficult to select such an interaction even for systems with a discrete variable (for example, spin or polarization); therefore, unneeded outcomes arise, which must be discarded. For example, the nonorthogonal resolution of the identity operator (44) can be realized by fusing a pair of photons into a single photon when they pass through a nonlinear crystal (parametric energy upconversion) and then detecting it by a photodetector.[11] However, because of the small nonlinear susceptibility a set of empty results appears and must be discarded.

6. CONCLUSION

Thus, measurements used in quantum teleportation have been considered in this paper from the standpoint of the general theory of quantum-mechanical measurements. It has been shown that the teleported state is completely specified by the resolution of the identity operator (positive operator-valued measure) generated by the respective instrument (quantum operation describing the change in the state of the system as a result of the measurement) in the state space of the system, so that there is actually no need to completely specify the instrument which gives the most complete description of the effects of the measurement procedure on the quantum system. A protocol for the quantum teleportation of a state of a system with a nondegenerate continuous spectrum based on nonorthogonal measurements has been proposed. In this protocol, as in all the other known protocols which ensure exact teleportation, an ideal EPR pair with singular correlations, which corresponds to an unnormalized wave function, must be used.[1)] The question of the possibility of the exact teleportation of a continuous quantum variable using physically realized (normalized) states remains open. Thus, there is still no known algorithm for the exact teleportation of a continuous variable for nonsingular EPR states.

As a final note, we thank K.A. Valiev for discussing the results obtained in this paper. This work was supported by the Russian Fund for Fundamental Research (Grant No. 99-02-18127) and Grant No. 02.04.5.2.40.T.50 from the ''Promising Technologies and Micro- and Nanoelectronics Devices'' Program.

*)E-mail: molotkov@issp.ac.ru

1)Generally speaking, the correct treatment of such states requires the enlistment of dressed Hilbert spaces. In our treatment the problem of the infinite norm of the EPR states used can be circumvented, because only the relative probabilities of a measurement result belonging to a particular set are significant for us.

[1] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).
[2] L. Vaidman, Phys. Rev. A **49**, 1473 (1994).
[3] S. L. Braunstein and H. J. Kimble, Phys. Rev. Lett. **80**, 869 (1998).
[4] M. Ozawa, J. Math. Phys. **34**, 5596 (1993).
[5] K. Kraus, *States, Effects and Operations*, Springer-Verlag, Berlin (1983).
[6] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, North-Holland, Amsterdam–New York (1982) [Russ. original, Nauka, Moscow (1980)].
[7] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton University, Princeton, NJ (1955).
[8] G. Lüders, Ann. Phys. (Leipzig) **8(6)**, 322 (1951).
[9] M. A. Nielsen and C. M. Caves, LANL E-print Archives http://xxx.lanl.gov/abs/quant-ph/9608001.
[10] S. N. Molotkov, Phys. Lett. A **245**, 339 (1998); LANL E-print Archives http://xxx.lanl.gov/abs/quant-ph/9805045.
[11] S. N. Molotkov, JETP Lett. **68**, 263 (1998); LANL E-print Archives http://xxx.lanl.gov/abs/quant-ph/9807013.
[12] P. van Loock, S. L. Braunstein, and H. J. Kimble, LANL E-print Archives http://xxx.lanl.gov/abs/quant-ph/9902030.

Translated by P. Shelnitz

Interaction of an atom with superstrong laser fields

A. V. Andreev*)

Department of Physics and International Laser Center, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 10 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 793–806 (September 1999)

A theory of atomic interaction with a superstrong laser field has been developed. The specific feature of the suggested theory is that its small parameter is the interaction between the atom and the solenoidal part of the external field, whereas its interaction with the potential part is accurately taken into account. It follows from the reported investigation that, in calculating the interaction of atoms with superstrong fields, one must abandon calculations of multipole moments of transitions between unperturbed atomic levels, and calculate instead the atomic response, which comprises multipole moments of all orders and depends on the instantaneous field magnitude. The results are compared with calculations based on the perturbation theory in terms of the interaction Hamiltonian. © *1999 American Institute of Physics.* [S1063-7761(99)00409-6]

1. INTRODUCTION

In recent years a lot of researchers' attention has been attracted to interaction of isolated atoms and molecules, and dense media as well (such as pressurized gases, plasmas, and liquids) with ultrashort laser pulses of high intensity. The response of a medium to a laser field of high intensity is highly nonlinear. This leads to generation of high harmonics, Raman components, or a quasicontinuum, i.e., the process leads to generation of waves of different frequencies.[1–6] The difference from quasi-stationary processes generating multiple frequencies is that the spectrum of the generated waves varies as a laser pulse propagates through a medium. The existing theories of an atom's interaction with a superstrong laser field (the theoretical approaches were reviewed in detail in Ref. 4) are usually based on treating the interatomic potential as a small parameter of the perturbation theory, since the Hamiltonian of interaction with the external field ceases to be a small parameter. Numerical and analytic calculations[4] have allowed the researchers to explain the basic features of effects observed in experiments.

This paper describes a theory of atomic interaction with a superstrong laser field, which has two distinctive features. First, the small parameter is the atomic interaction with the solenoidal part of the external field, and the interaction with the potential part is described in an exact form. Second, the reported investigation has led to a conclusion that one must abandon calculations of multipole moments of electronic transitions between unperturbed states of an atom in studying its interaction with laser fields, and calculate instead the atomic response at a certain frequency with due account of multipole moments of all orders up to infinity as a function of the instantaneous laser field magnitude. It follows from the calculations that the atomic response at the frequency of the nth harmonic of the laser field in this case is proportional to the corresponding power of the field only in the weak-field limit. When the external field strength becomes comparable to that of the intratomic field, the response amplitude saturates and then drops with the external wave intensity. This explains the multiplicity of harmonic generation, i.e., the fact that amplitudes of high-order harmonics in the atomic response become comparable.

The basic equations of the suggested theory are given in Sec. 2. Then the specific features of atomic interaction with superstrong laser fields are discussed. Section 4 describes relativistic corrections to the Hamiltonian of atomic interaction with a superstrong laser field. Section 5 derives an operator equation for the current density and analyzes the effect of gradient forces.

2. EQUATION DESCRIBING ATOMIC INTERACTION WITH A SUPERSTRONG LASER FIELD

2.1. General case

Schrödinger's equation for an atom interacting with a transverse electromagnetic field in the nonrelativistic approximation has the form

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi, \tag{1}$$

where

$$H = \frac{1}{2m}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)^2 + U_0. \tag{2}$$

There are two basic approaches to the problem of atomic interaction with electromagnetic field. When the field strength in the incident electromagnetic wave is much smaller than the intratomic field strength, the perturbation theory is widely used. The basic assumption in this case is that the free-atom Hamiltonian

$$H_0 = \frac{p^2}{2m} + U_0 \tag{3}$$

is much larger than the interaction Hamiltonian

$$H_{\rm int}=-\frac{e}{2mc}(\mathbf{p}\cdot\mathbf{A}+\mathbf{A}\cdot\mathbf{p})+\frac{e^2}{2mc^2}\mathbf{A}^2, \tag{4}$$

and a solution of Eq. (1) is sought in the form of an expansion in powers of the external field.

In a strong external field, the intratomic potential U_0 in Eq. (2) is treated as a perturbation.

Both of these approximations have their specific application domains. In the problem under discussion, however, one can find another small parameter which is independent of the ratio between the strengths of the external and intratomic fields, so this approach allows one to track changes in an atom's reaction to external field when the latter increases from very weak to superstrong. This parameter is the ratio between the potential and solenoidal parts of the external field.

To separate the potential part of the vector potential $\mathbf{A}(\mathbf{r},t)$, we use the identity

$$\mathrm{grad}(\mathbf{A}\cdot\mathbf{r})=\mathbf{r}\times\mathrm{curl}\,\mathbf{A}+(\mathbf{r}\nabla)\mathbf{A}+\mathbf{A}.$$

Hence

$$\mathbf{A}(\mathbf{r},t)=\mathbf{A}_1(\mathbf{r},t)+\mathbf{A}_2(\mathbf{r},t), \tag{5}$$

where

$$\mathbf{A}_1=\mathrm{grad}(\mathbf{A}\cdot\mathbf{r}),\quad \mathbf{A}_2=\mathbf{H}\times\mathbf{r}-(\mathbf{r}\nabla)\mathbf{A}.$$

Thus, with due account of Eq. (5), the wave equation (1) can be rewritten as

$$i\hbar\frac{\partial\psi}{\partial t}=\exp\left[i\frac{e}{\hbar c}\chi(\mathbf{r},t)\right]H(\mathbf{A}_2)\exp\left[-i\frac{e}{\hbar c}\chi(\mathbf{r},t)\right]\psi, \tag{6}$$

where

$$\chi(\mathbf{r},t)=\mathbf{A}(\mathbf{r},t)\cdot\mathbf{r}.$$

Equation (6) has the form

$$i\hbar\frac{\partial\psi}{\partial t}=V(H_0+H_{\rm int}(\mathbf{A}_2))V^{-1}\psi, \tag{7}$$

where

$$V=\exp\left[i\frac{e}{\hbar c}\chi(\mathbf{r},t)\right].$$

Note that

$$VH_0V^{-1}=\frac{1}{2m}\left(\mathbf{p}-\frac{e}{c}\mathbf{A}_1\right)^2+U_0,$$

$$VH_{\rm int}(\mathbf{A}_2)V^{-1}=-\frac{e}{2mc}(\mathbf{p}\cdot\mathbf{A}_2+\mathbf{A}_2\cdot\mathbf{p})+\frac{e^2}{2mc^2}(\mathbf{A}^2-\mathbf{A}_1^2).$$

It follows directly from Eq. (5) that at $\lambda\gg a$, i.e., when the laser wavelength λ is much larger than the amplitude a of electron oscillations in the external field,

$$|\mathbf{A}_1|\gg|\mathbf{A}_2|.$$

We can therefore seek a solution of Eq. (7) using an iterative method:

$$i\hbar\frac{\partial\psi_0}{\partial t}=VH_0V^{-1}\psi_0,$$

$$i\frac{\partial\psi_1}{\partial t}=VH_0V^{-1}\psi_1+VH_{\rm int}(\mathbf{A}_2)V^{-1}\psi_0,\ldots \tag{8}$$

Let us focus attention on the zero-order approximation of Eq. (8). Expand the wave function $\psi_0(\mathbf{r},t)$ in terms of the eigenfunctions of both the discrete spectrum and the continuum generated by the intratomic Hamiltonian H_0:

$$\psi_0(\mathbf{r},t)=\sum_n a_n(t)u_n(\mathbf{r})+\int d\mathbf{k}a_{\mathbf{k}}(t)u_{\mathbf{k}}(\mathbf{r}).$$

By substituting this expansion in Eq. (8), we obtain

$$\frac{da_n}{dt}=-i\sum_{m,l}\left(\exp\left[i\frac{e\mathbf{A}\cdot\mathbf{r}}{\hbar c}\right]\right)_{nm}\omega_m\left(\exp\left[-i\frac{e\mathbf{A}\cdot\mathbf{r}}{\hbar c}\right]\right)_{ml}a_l. \tag{9}$$

Note that the sum in Eq. (9) also includes integration over the continuum states, which is not reflected in the formula for simplicity.

Let us investigate evolution of the wave function of atomic electrons under an intense monochromatic laser pulse of the form

$$\mathbf{A}(\mathbf{r},t)=\mathbf{A}_0(\boldsymbol{\rho},t)\sin(\omega t-\kappa z), \tag{10}$$

where $\boldsymbol{\rho}$ is the transverse coordinate in the beam cross section. In interpreting the equations to follow, it is more convenient to use the electric field amplitude $\mathbf{E}_0$ of the laser pulse instead of the vector potential $\mathbf{A}_0$. Recall, therefore, that when

$$\left|\frac{\partial A_0}{\partial t}\right|\ll\omega|A_0|,$$

one can set

$$\mu=\frac{e\mathbf{A}_0\cdot\mathbf{r}}{\hbar c}\approx-\frac{e\mathbf{E}_0\cdot\mathbf{r}}{\hbar\omega}.$$

Thus, a power-series expansion of the exponential function in Eq. (9) when $e\mathbf{E}_0\cdot\mathbf{r}\ll\hbar\omega$ yields the standard perturbation expansion in powers of the field strength. However, if one uses the generating functions for the Bessel equation

$$\sin(\mu\sin\theta)=2\sum_{n=0}^{\infty}J_{2n+1}(\mu)\sin((2n+1)\theta),$$

$$\cos(\mu\sin\theta)=J_0(\mu)+2\sum_{n=1}^{\infty}J_{2n}(\mu)\cos(2n\theta), \tag{11}$$

one can derive from Eq. (9) an explicit expression for the atomic response at the nth harmonic frequency that includes all powers of the field strength. For $e\mathbf{E}_0\cdot\mathbf{r}\ll\hbar\omega$, the largest term in the expansion of the Bessel function is the first, which is proportional to the corresponding power of the field strength. The Bessel function, however, decreases with increasing argument, i.e., field amplitude. Therefore, the atomic response at the frequency of an incident wave harmonic is a nonmonotonic function of the field amplitude if the inequality given above does not hold. The atomic re-

sponse increases initially and then decreases, starting with a specific field amplitude that depends on the ordere of the harmonic. This fact alone indicates a substantial difference between Eq. (9) and the various versions of the conventional perturbation theory, which are based on a power-series expansion in terms of the incident wave amplitude.

2.2. Perturbation theory

Let us analyze in more detail the interaction between atoms and superstrong laser fields. As noted above, in the conventional perturbation theory an atomic response at the nth harmonic frequency is proportional to the corresponding power of the field amplitude, and the polarizability is proportional to the $(n+1)$th power of the dipole moment. The approach suggested in this paper treats an atomic response as a function of matrix elements calculated on the basis of the Bessel function. Let us analyze the impact of this modification of the mathematical model on the description of physical effects.

Let us start with the simplest case. Assume that the atom was in its ground state before the arrival of the laser pulse $[a_n(t=0)=\delta_{n0}]$, and the changes in the populations of excited states during the pulse duration are negligible in comparison with that of the ground state.

In this case, we obtain from Eq. (9)

$$\begin{aligned}\frac{da_n}{dt}=&-i(\omega_n+\Delta\omega_n)a_n-2i\sum_m\left(\sum_{k=1}^{\infty}[J_{2k}\cos(2k\theta)\right.\\&\left.+iJ_{2k+1}\sin(2k+1)\theta]\right)_{nm}\omega_m(J_0(\mu))_{m0}\\&-2i\sum_m(J_0(\mu))_{nm}\omega_m\left(\sum_{k=1}^{\infty}[J_{2k}\cos(2k\theta)\right.\\&\left.-iJ_{2k+1}\sin(2k+1)\theta]\right)_{m0}-4i\\&\times\sum_m\left(\sum_{k=1}^{\infty}[J_{2k}\cos(2k\theta)+iJ_{2k+1}\sin(2k+1)\theta]\right)_{nm}\\&\times\omega_m\left(\sum_{k=1}^{\infty}[J_{2k}\cos(2k\theta)-iJ_{2k+1}\sin(2k+1)\theta]\right)_{m0},\end{aligned}\tag{12}$$

where $\theta=\omega t-\kappa z$, and the Stark shift of the nth level is

$$\Delta\omega_n=\sum_m(J_0(\mu))_{nm}\omega_m(J_0(\mu))_{mn}-\omega_n.\tag{13}$$

Assuming that the incident wave has a linear polarization aligned with the x-axis, we have $\mu=eA_0x/\hbar c$ in Eqs. (9) and (12). In the optical range, the wavelength is much greater than sizes of atomic shells, $\lambda\gg a_0$, so we can set $z=0$ in calculating matrix elements and neglect the dependence of A_0 on the transverse coordinates, since the laser beam transverse size is of order of the wavelength even at the lens focal point. Consequently, the matrix elements in Eq. (12) have the form $[J_n(\alpha x)]_{kl}$.

2.3. Resonant case

Suppose that one of the field harmonics is resonant with an atomic transition, i.e.,

$$n_0\omega\approx\omega_{k0}+\Delta\omega_{k0}.$$

In this case, we obtain from Eq. (12)

$$\begin{aligned}a_k(t)\approx&\frac{1}{\omega_{k0}+\Delta\omega_{k0}-n_0\omega}\sum_l\left\{\sum_{m=0}^{n_0}(-1)^{n_0-m}\right.\\&\times(J_{n_0-m}(\alpha x))_{kl}\omega_l(J_m(\alpha x))_{l0}+(-1)^{n_0}\\&\times\sum_{m=1}^{\infty}(J_{n_0+m}(\alpha x))_{kl}\omega_l(J_m(\alpha x))_{l0}\\&\left.+\sum_{m=1}^{\infty}(J_m(\alpha x))_{kl}\omega_l(J_{n_0+m}(\alpha x))_{l0}\right\}.\end{aligned}\tag{14}$$

On the other hand, using the power-series expansion of the exponential function in accordance with the conventional perturbation theory, we obtain

$$\begin{aligned}a_k^{(P)}(t)\approx&\frac{1}{\omega_{k0}-n_0\omega}\sum_l\left\{\alpha^{n_0}\sum_{m=0}^{n_0}(-1)^{n_0-m}\right.\\&\left.\times\frac{(x^{n_0-m})_{kl}\omega_l(x^m)_{l0}}{2^{n_0}(n_0-m)!m!}+\ldots\right\}.\end{aligned}\tag{15}$$

The comparison between Eqs. (14) and (15) demonstrates that the difference between the results obtained using these two approaches is caused by the different forms of the matrix elements in the expressions for the level population amplitudes. Let us demonstrate that Eqs. (14) and (15) predict different physical effects.

3. FEATURES OF INTERACTION BETWEEN AN ATOM AND A SUPERSTRONG LASER FIELD

3.1. Harmonic oscillator

The linear harmonic oscillator is a natural first example of a system interacting with a laser pulse. The matrix elements of transitions between the levels of the harmonic oscillator are well known:

$$(x)_{n,n-1}=\sqrt{\frac{n\hbar}{2m\omega_0}}.\tag{16}$$

Let us assume first that the laser pulse carrier frequency coincides with that of the transitions in the harmonic oscillator, i.e., $\omega\approx\omega_0$. Using the expressions for the wave functions of the harmonic oscillator, one can easily obtain

$$(J_1(\alpha x))_{10}=\frac{\beta}{2}\exp\left(-\frac{\beta^2}{4}\right)\left[I_0\left(\frac{\beta^2}{4}\right)-I_1\left(\frac{\beta^2}{4}\right)\right],\tag{17}$$

where $I_n(z)$ is the modified Bessel function,

$$\beta=\alpha\sqrt{\frac{\hbar}{2m\omega_0}}=\frac{eA_0}{\hbar c}a_0=\frac{eE_0a_0}{\hbar\omega}. \tag{18}$$

Substituting Eqs. (16) and (17) into (14) and (15), we obtain

$$\frac{a_1}{a_1^{(P)}}=\exp\left(-\frac{\beta^2}{4}\right)\left[I_0\left(\frac{\beta^2}{4}\right)-I_1\left(\frac{\beta^2}{4}\right)\right]. \tag{19}$$

The parameter β in Eq. (19) is the ratio between the force of the external field driving the electron, $F_e=eE_0$, and the force due to the intratomic potential:

$$F_{\text{at}}=|\partial U_0/\partial x|\approx\hbar\omega_0/a_0.$$

Thus, ratio (19) equals unity for a field of a moderate intensity ($\beta\ll1$). This ratio, however, decreases with the external field amplitude, i.e., the perturbation theory overestimates the population of excited atomic states due to interaction with a superstrong laser field.

3.2. Forbidden transitions

Let the external field be resonant with the $0\leftrightarrow3$ transition of the harmonic oscillator, i.e., $\omega\approx3\omega_0$. The dipole matrix element for this transition is identically zero, $(x)_{30}\equiv0$. On the other hand, using the harmonic oscillator wave functions, we obtain

$$(J_1(\alpha x))_{30}=\frac{\beta^3}{2\sqrt{6}}\exp\left(-\frac{\beta^2}{4}\right)\left[\left(1+\frac{2}{\beta^2}\right)I_1\left(\frac{\beta^2}{4}\right)-I_0\left(\frac{\beta^2}{4}\right)\right].$$

Thus, this example clearly demonstrates that dipole-forbidden transitions in atoms and molecules interact with resonant superstrong laser fields, and at $\beta\approx1$ the populations of the corresponding higher levels are comparable to those of levels connected to the ground state via dipole-allowed transitions.

3.3. Hydrogen atom

Now let us consider the hydrogen atom as a more realistic quantum mechanical model. Let the nth harmonic of the driving field be in the resonance with the transition $(n=1,l=0,m=0)\leftrightarrow(n=2,l=1,m=0)$. It follows from symmetry considerations that the harmonic number must be odd, i.e., $(2k+1)\omega=\omega_{21}$. Nonetheless, in order not to complicate the formulas, we will use the label n. Using the atomic wave functions of hydrogen, we easily obtain

$$(J_n(\alpha x))_{21}=\frac{1}{\sqrt{2}}\left(\frac{2}{3}\right)^4\beta^n \times\frac{3(1+n\sqrt{1+\beta^2})+n(1+\beta^2)(n+\sqrt{1+\beta^2})}{(1+\beta^2)^{5/2}(1+\sqrt{1+\beta^2})^n}. \tag{20}$$

For a given transition, the parameter β, whose physical sense is the same as in the previous sections, is given by

$$\beta=\frac{eA_0}{\hbar c}\frac{2a_B}{3},$$

where a_B is the Bohr radius. Expression (20) has the following asymptotic behavior:

$$(J_n(\alpha x))_{21}=\frac{1}{\sqrt{2}}\left(\frac{2}{3}\right)^4\begin{cases}(\beta/2)^n(n+1)(n+3), & \beta\ll1,\\ n/\beta^2, & \beta\gg1.\end{cases} \tag{21}$$

Comparing these with the matrix elements calculated on the basis of expansion terms of the exponential function proportional to the nth power of the field strength, we have

$$\left(\frac{\alpha^n x^n}{2^n n!}\right)_{21}=\frac{1}{\sqrt{2}}\left(\frac{2}{3}\right)^4\left(\frac{\beta}{2}\right)^n(n+1)(n+3). \tag{22}$$

It is clear that the matrix elements calculated by the two different methods coincide in the low-field limit. In the superstrong field, however, there is a significant difference between the results. For example, in the low-field limit the ratio $(J_n)_{21}/(J_1)_{21}$ is a power-law function of the field amplitude ($\propto\beta^{n-1}$), whereas in the high-field limit this ratio no longer depends on the field amplitude, and the matrix elements decrease with increasing field strength.

3.4. Ionization in a superstrong laser field

The suggested approach leads to fundamentally different results concerning ionization of atoms by superstrong laser fields. This conclusion can be drawn from general considerations. The matrix elements of the n-photon transition between the bound atomic states and continuum for odd n take the general form

$$\int\frac{e^{ikr}}{r}x^n e^{-r/a_0}\cos\theta\,dV.$$

One can easily demonstrate that they decrease rapidly with increasing kinetic energy of ionized electrons, $E_k=\hbar^2k^2/2m$. On the other hand, owing to the fast oscillating character of the Bessel functions, matrix elements involving these functions do not decrease with E_k when $\alpha a_0\gg1$:

$$\int\frac{e^{ikr}}{r}J_n(ar\cos\theta)e^{-r/a_0}\cos\theta\,dV\approx\int\sqrt{\frac{|\cos\theta|}{2\pi\alpha r}}\frac{1}{r} \times\exp\left[i\left(kr-\alpha r\cos\theta-\frac{\pi(2n+1)}{4}\right)-\frac{r}{a_0}\right]dV.$$

The latter equation clearly shows that these matrix elements have maxima if the condition $k\approx\alpha$ is satisfied. This condition and energy conservation, $E_k=E_0+n\hbar\omega$, determine the energy distribution of ionized electrons. Thus, for a laser beam of a nonuniform cross section, a set of harmonics will be generated, and the highest harmonic number is determined by the field intensity on the beam axis.

3.5. Generation of even harmonics

Up to this point, we have not taken into consideration the coordinate dependence of the vector potential in the calculations of matrix elements, i.e., we have assumed that $\mathbf{A}_0(\mathbf{r},t)=\mathbf{A}_0(\mathbf{r}_0,t)$, where $\mathbf{r}_0$ is the location of the atomic nucleus. In this case, the selection rules for the matrix elements in the basis of functions $J_n(\alpha x)$ and x^n are identical

owing to their identical symmetry properties. However, it will be demonstrated in what follows that the coordinate dependence of the vector potential becomes an important factor in the focused laser beams. In the general case, the matrix elements take the form

$$\left(J_n\left(\frac{e(x-x_0)}{\hbar c}(A_0(\mathbf{r}_0)+((\mathbf{r}-\mathbf{r}_0)\nabla)A_0|_{\mathbf{r}=\mathbf{r}_0}+\ldots)\right) \times \exp(in\kappa z)\right)_{kl}. \tag{23}$$

Even taking account of just the first-order term changes the symmetry properties of the Bessel functions; as a result, the selection rules are modified. In the limit of high-intensity fields, matrix elements (23) will be different from the matrix elements calculated using the expansion in powers of the field amplitude, even if the coordinate dependence of the field envelope is taken into account. This shows up most clearly in the relationship between the intensities of even and odd harmonics in the field generated by the atom.

4. RELATIVISTIC CORRECTIONS TO HAMILTONIAN OF INTERACTION WITH SUPERSTRONG LASER FIELD

4.1. Relativistic corrections

In superstrong laser fields with intensities much greater than that of the intratomic field, relativistic effects become important, so we now dwell in more detail on calculations of relativistic corrections to Hamiltonian (2).

The interaction Hamiltonian of an atom and electromagnetic field can be expressed in the second-quantization representation as

$$H=\int \Psi^+[\boldsymbol{\alpha}\cdot(c\mathbf{p}-e\mathbf{A})\Psi+e\varphi\Psi+mc^2\beta\Psi]dV + \int\left[2\pi c^2\mathbf{B}^2+\frac{1}{8\pi}(\operatorname{curl}\mathbf{A})^2-c\mathbf{B}\cdot\operatorname{grad}\varphi\right]dV, \tag{24}$$

where $\boldsymbol{\alpha}$ and β are Dirac matrices. In the second-quantization representation, the wave functions $\Psi(\mathbf{r},t)$ and $\Psi^+(\mathbf{r},t)$, the vector potential of external electromagnetic field $\mathbf{A}(\mathbf{r},t)$, and the generalized momentum $\mathbf{B}(\mathbf{r},t)$ canonically conjugate to the vector potential satisfy the commutation relations

$$\begin{aligned}
&[\Psi(\mathbf{r},t),\Psi(\mathbf{r}',t)]_-=[\Psi^+(\mathbf{r},t),\Psi^+(\mathbf{r}',t)]_-=0,\\
&[\Psi(\mathbf{r},t),\Psi^+(\mathbf{r}',t)]_-=\delta(\mathbf{r}-\mathbf{r}'),\\
&[A_\alpha(\mathbf{r},t),A_\beta(\mathbf{r}',t)]_-=[B_\alpha(\mathbf{r},t),B_\beta(\mathbf{r}',t)]_-=0,\\
&[A_\alpha(\mathbf{r},t),B_\beta(\mathbf{r}',t)]_-=i\hbar\,\delta_{\alpha\beta}\delta(\mathbf{r}-\mathbf{r}'),
\end{aligned} \tag{25}$$

which lead to the Dirac wave equation and well known equations for the electromagnetic field:

$$i\hbar\frac{\partial\Psi}{\partial t}=c\boldsymbol{\alpha}\cdot\left(\mathbf{p}-\frac{e}{c}\mathbf{A}\right)\Psi+e\varphi\Psi+\beta mc^2\Psi, \tag{26a}$$

$$\frac{\partial\mathbf{A}}{\partial t}=4\pi c^2\mathbf{B}-\mathrm{c}\operatorname{grad}\varphi, \tag{26b}$$

$$\frac{\partial\mathbf{B}}{\partial t}=-\frac{1}{4\pi}\operatorname{curl}\operatorname{curl}\mathbf{A}-e\Psi^+\boldsymbol{\alpha}\Psi. \tag{26c}$$

Introducing the notation

$$\mathbf{P}=\mathbf{p}-\frac{e}{c}\mathbf{A},\quad U=e\varphi$$

and expressing the wave function in the form

$$\Psi(\mathbf{r},t)=\begin{pmatrix}\xi\\ \eta\end{pmatrix}\exp\left[-i\frac{mc^2}{\hbar}t\right],$$

we easily transform Eq. (26a) to

$$\left(i\hbar\frac{\partial}{\partial t}-U\right)\xi=c\boldsymbol{\sigma}\cdot\mathbf{P}_\eta,\quad \left(i\hbar\frac{\partial}{\partial t}-U+2mc^2\right)\eta=c\boldsymbol{\sigma}\cdot\mathbf{P}\xi. \tag{27}$$

From the second equation in (27), we obtain

$$\eta=\left(2mc^2-U+i\hbar\frac{\partial}{\partial t}\right)^{-1}c\boldsymbol{\sigma}\cdot\mathbf{P}\xi = \frac{1}{2mc^2-U}\sum_{n=0}^{\infty}\left(i\hbar\frac{\partial}{\partial t}\frac{1}{2mc^2-U}\right)^n c\boldsymbol{\sigma}\cdot\mathbf{P}\xi. \tag{28}$$

When $U,E\ll mc^2$, one can take into account only the two lowest-order terms on the right-hand side of Eq. (28). As a result, the wave equation for an electron in an electromagnetic field with the vector and scalar potentials $\mathbf{A}(\mathbf{r},t)$ and $\varphi(\mathbf{r},t)$ has the form

$$\begin{aligned}
i\hbar\frac{\partial\psi}{\partial t}=&\left(\frac{1}{2m_1}\left(\mathbf{p}-\frac{e}{c}\mathbf{A}\right)^2+U\right)\psi-\frac{e\hbar}{2m_1c}\boldsymbol{\sigma}\cdot\mathbf{H}\psi\\
&+\frac{\hbar c^2}{(2mc^2-U)^2}\boldsymbol{\sigma}\cdot\left[\frac{e}{c}\frac{\partial\mathbf{A}}{\partial t}+\nabla U,\mathbf{p}-\frac{e}{c}\mathbf{A}\right]\psi\\
&+\frac{\hbar^2c^2}{2}\operatorname{div}\left(\frac{(e/c)(\partial\mathbf{A}/\partial t)+\nabla U}{(2mc^2-U)^2}\right)\psi+i\frac{\hbar^2e}{2c}\\
&\times\boldsymbol{\sigma}\cdot\operatorname{curl}\left(\frac{c^2}{(2mc^2-U)^2}\frac{\partial\mathbf{A}}{\partial t}\right)\psi\\
&+\hbar\boldsymbol{\sigma}\cdot\left[\nabla\frac{1}{2m_1},\mathbf{p}-\frac{e}{c}\mathbf{A}\right]\psi,
\end{aligned} \tag{29}$$

where

$$\psi(\mathbf{r},t)=\left(1+\boldsymbol{\sigma}\cdot\mathbf{P}\frac{c^2}{(2mc^2-U)^2}\boldsymbol{\sigma}\cdot\mathbf{P}\right)\xi(\mathbf{r},t),$$

$$\frac{1}{m_1}=\frac{1}{m}\left(1-\left(\frac{U}{2mc^2-U}\right)^2\right),\quad \mathbf{H}=\operatorname{curl}\mathbf{A}. \tag{30}$$

The continuity equation

$$\frac{\partial\rho}{\partial t}+\operatorname{div}\mathbf{J}=0$$

follows directly from the wave equation (29), where the current density operator has the form

$$\mathbf{J}=\frac{i\hbar e}{2m_1}(\nabla\psi^+\psi-\psi^+\nabla\psi)-\frac{e^2}{m_1c}\psi^+\mathbf{A}\psi$$
$$-\frac{\hbar e^2c^2}{(2mc^2-U)^2}[\mathbf{E},\psi^+\boldsymbol{\sigma}\psi]. \quad (31)$$

The general scheme described in Sec. 2 can also be applied to Eq. (29). The interaction Hamiltonian $H_{\rm int}(\mathbf{A}_2)$ in this case has the form

$$H_{\rm int}(\mathbf{A}_2)=-\frac{e}{2mc}(\mathbf{p}\cdot\mathbf{A}_2+\mathbf{A}_2\cdot\mathbf{p})+\frac{e^2}{2mc^2}\mathbf{A}_2^2$$
$$+H_{\rm rel}(\mathbf{A}_2)+H_{\rm spin}(\mathbf{A}_2).$$

This operator takes into account the effects of the nonlocal nature of interaction due to the nonpotential parts of field $\mathbf{A}_2$, changes in the electron mass, and spin effects.

4.2. Ionization stabilization

The Hamiltonian in wave equation (29) is different from the conventional Hamiltonian that takes into account quadratic relativistic corrections.[7] First of all, the corrections due to the spin-orbit and contact interactions take into account an interaction not only with the static field of the nucleus, but also with the transverse electromagnetic field. In order to clarify the resulting differences, recall that we wrote down the single-electron Hamiltonian (24) in the interest of brevity. In the case of an atom, summation over all charges should be performed on the right-hand side of Eq. (24). If the atom is driven by an external electromagnetic wave, the vector and scalar potentials incorporate the intratomic field, external field, and the field generated by the atomic electrons:

$$\mathbf{A}=\mathbf{A}_0+\mathbf{A}_e+\mathbf{A}',\quad \varphi=\varphi_0+\varphi_e+\varphi'.$$

If we neglect motion of the nucleus in the external field, the intratomic potential can be treated as static $[\mathbf{A}_0=0,\varphi_0=\varphi(\mathbf{r})]$. The external field can be deemed transverse $(\operatorname{div}\mathbf{A}_e=0,\varphi_e=0)$. In this case, the Hamiltonian of contact interaction due to the fourth term on the right-hand side of Eq. (29) at $U\ll mc^2$ has the form

$$H_c=\frac{e\hbar^2}{8m^2c^2}\operatorname{div}\left(\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}+\nabla\varphi\right)=\frac{e\hbar^2}{8m^2c^2}\Delta\varphi_0$$
$$-\frac{e\hbar^2}{8m^2c^2}\operatorname{div}\mathbf{E}=\frac{\pi e^2\hbar^2Z}{2m^2c^2}\delta(\mathbf{r})-\frac{\pi e^2\hbar^2}{2m^2c^2}|\psi|^2. \quad (32)$$

Similarly, the spin-orbit Hamiltonian acquires, in addition to the term due to the electron motion in the atomic potential, a term caused by their motion in the external wave field.

It is noteworthy, however, that under conditions of a high ionization degree, a fundamental role is played by the modified Hamiltonian (32) of contact interaction. The contribution of this Hamiltonian, which is proportional to the wave function modulus squared, demonstrates that a decrease in the electron density leads to a higher energy of the atom interacting with electromagnetic field. This can be one of the causes of the ionization stabilization.[6]

4.3. Mass corrections

The second feature of the Hamiltonian in Eq. (29) is that it takes into account the effect of external-field potential $\varphi_e(\mathbf{r},t)$ (through relativistic corrections to the mass). Note that in deriving Eq. (29), we took into account the time dependence of the scalar potential

$$U(\mathbf{r},t)=U_0(\mathbf{r})+e\varphi_e(\mathbf{r},t),$$

so a gradient gauge transformation of the wave function

$$\Psi(\mathbf{r},t)=\Psi'(\mathbf{r},t)\exp\left(i\frac{e}{\hbar c}\mathbf{A}(\mathbf{r},t)\cdot\mathbf{r}\right),$$

transforms the scalar and vector potentials in Eq. (29) in accordance with the well-known relations

$$\mathbf{A}'(\mathbf{r},t)=\mathbf{A}_2(\mathbf{r},t),\quad \varphi_e'(\mathbf{r},t)=-e\mathbf{E}\cdot\mathbf{r}\approx U_p. \quad (33)$$

Substituting Eq. (33) into (30), we see that relativistic mass corrections are important when the ponderomotive potential U_p approaches mc^2.

5. GRADIENT FORCES

In the previous sections, we have used the quantum mechanical approach based on calculation of wave functions. In many cases, however, it is more convenient to utilize the operator approach, because the form of the resulting equations is more similar to the classical equations. We have already derived the equation for the charge density. Using Hamiltonian (24) and commutation relations (25), it is not difficult to obtain an equation for the current density. If relativistic corrections are neglected, the expression for the current density operator can be written in the form

$$\mathbf{J}=\frac{i\hbar e}{2m}(\nabla\psi^+\psi-\psi^+\nabla\psi)-\frac{e^2}{mc}\psi^+\mathbf{A}\psi=\mathbf{j}-\frac{e^2}{mc}\psi^+\mathbf{A}\psi.$$

The equations for $\mathbf{J}$ have the form

$$\frac{\partial J_\alpha}{\partial t}-\frac{e}{mc}\mathbf{A}\cdot\nabla J_\alpha=-\frac{i}{\hbar}[j_\alpha,H_a]_-+\frac{e}{m}E_\alpha\rho$$
$$+\frac{e}{mc}[\mathbf{J}\operatorname{curl}\mathbf{A}]_\alpha+\left(\frac{e}{mc}\right)^2A_\beta\nabla_\beta(A_\alpha\rho)$$
$$+\frac{e}{mc}\nabla_\beta(A_\alpha J_\beta), \quad (34)$$

where $\alpha,\beta=x,y,z$, and the convention of summation over repeated indices is adopted in Eq. (34). The commutator of $\mathbf{j}$ and the Hamiltonian H_a of intratomic interaction takes the form

$$[j_\alpha,H_a]_-=-\frac{i\hbar}{m}\rho\nabla_\alpha U+\frac{i\hbar^3}{4m^2}\nabla_\alpha(\Delta\rho)$$
$$-\frac{ie\hbar^3}{2m^2}\nabla_\beta(\nabla_\alpha\psi^+\nabla_\beta\psi+\nabla_\beta\psi^+\nabla_\alpha\psi). \quad (35)$$

Substitution of Eq. (35) in (34) yields an equation similar to the classical one, with the exception of gradient terms. After integration over a volume much larger that the amplitude of electron oscillations, some gradient terms cancel out.

The set of equations for the charge and current densities is presented in a more graphic form in the case of the Coulomb gauge of the electromagnetic field ($\operatorname{div}\mathbf{A}=0$):

$$\frac{\partial\rho}{\partial t}-\frac{e}{mc}\mathbf{A}\cdot\nabla\rho=-\operatorname{div}\mathbf{j},$$

$$\frac{\partial j_\alpha}{\partial t}-\frac{e}{mc}\mathbf{A}\cdot\nabla j_\alpha=-\frac{i}{\hbar}[j_\alpha,H_a]_-+\frac{e}{mc}j_\beta\nabla_\alpha(A_\beta)-\frac{1}{2}\left(\frac{e}{mc}\right)^2\rho\nabla_\alpha A^2. \quad (36)$$

The left-hand sides in Eq. (36) indicate that the time dependences of the charge and current densities are controlled by the relations

$$\rho(\mathbf{r},t)=\rho(\mathbf{r}-\mathbf{r}_0(t),t),\quad \mathbf{j}(\mathbf{r},t)=\mathbf{j}(\mathbf{r}-\mathbf{r}_0(t),t),$$

where

$$\frac{d\mathbf{r}_0}{dt}=\frac{e}{mc}\mathbf{A}.$$

Thus, their Fourier transforms

$$\rho(\mathbf{r},t)=\int d\mathbf{k}\rho(\mathbf{k},t)\exp[i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_0(t))]$$

again contain exponential functions of the vector potential, so we can again use expansion (11) in terms of Bessel functions.

The first gradient term on the right-hand side of Eq. (36) for the current density relates the various components of the polarization current density. Consequently, the field generated by an atom acted upon by an intense linearly polarized wave has elliptical polarization in the general case. The last term in this equation yields a gradient force acting on an electron in a nonuniform field. It equals the derivative of the ponderomotive potential

$$U_p=\frac{e^2A_0^2}{2mc^2}\approx\frac{e^2E_0^2}{2m\omega^2}.$$

At the focus of a diffraction-limited beam,

$$\left|\frac{\partial U_p}{\partial x}\right|\approx U_p/\lambda\,.$$

The force acting on the external electron of the atom is $\hbar\omega/a_B$. Noting that the ponderomotive potential is related to the field intensity by $U_p\,(\mathrm{eV})=10^{-13}I\,(\mathrm{W/cm^2})$, we find that

$$\left|\frac{\partial U_p}{\partial x}\right|\approx\hbar\omega/a_B$$

at $I=10^{17}\,\mathrm{W/cm^2}$ for $\hbar\omega=1-2$ eV and $\lambda=0.5-1\ \mu$m. At this power density, the external force driving the electron equals the intratomic force. Thus, the gradient force will generate a distribution of dissociated electrons that is nonuniform in the laser beam cross section.

6. CONCLUSIONS

The reported investigation has demonstrated that under laser fields of moderate intensity, the atomic response to the external field is essentially perfectly described by calculations based on the perturbation theory in terms of the interaction Hamiltonian. Nonetheless, the atomic response to an external field of strength comparable to that of the intratomic field is radically different for several reasons. First, the amplitude of the nth eigenstate of the unperturbed Hamiltonian in the superposition state

$$\psi(\mathbf{r},t)=\sum_n a_n(t)u_n(\mathbf{r})+\int d\mathbf{k}\,a_{\mathbf{k}}(t)u_{\mathbf{k}}(\mathbf{r}),$$

generated by the external wave is no longer determined by the dipole moment of the transition from the ground state (or other high-population states). Populations of states connected to the ground state via dipole-forbidden transitions can be greater than those of states to which dipole transitions are allowed. Second, atomic interaction with superstrong laser fields is highly nonlinear and demonstrates saturation. The response at the frequency of the nth harmonic is a power-law function of the field strength only if the latter is much less than that of the intratomic field. This equalizes amplitudes of different eigenstates in the superposition of different atomic states, and thereby leads to smaller differences between harmonic amplitudes in the induced field. If $\alpha a_0\gg 1$, i.e., we have a superstrong laser field, the amplitudes of ionized states with energies $E_k=E_0+n\hbar\omega$ no longer depend on the number n, but are controlled by the phase relations among them, with $\varphi_n\approx\pi(2n+1)/4$. Finally, the gradient forces due to the atom-field interaction on a distance scale of the order of the wavelength are comparable to the atomic forces when the external field strength is of the order of the intratomic field strength. As a result, the polarization response of an atom in a tightly focused beam is significantly different from its response to a plane optical wave with the same field strength. This effect provides additional tools for control of the spectrum of the resulting radiation.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 99-02-16093) and the Russian Universities program.

*)E-mail: andreev@srl.ilc.msu.su

[1] B. Luther-Davis, E. G. Gamalii, Ya. Wang, A. V. Rohde, and V. T. Tikhonchuk, Kvantovaya Électronika **19**, 317 (1992) [Sov. J. Quantum Electron. **22**, 289 (1992)].
[2] N. B. Delone and V. P. Krainov, *Multiphoton Process in Atoms*, Springer-Verlag, Berlin (1994).
[3] M. V. Fedorov, *Atomic and Free Electrons in a Strong Light Field*, World Scientific, Singapore, River Edge, NJ (1997).
[4] V. T. Platonenko and V. V. Strelkov, Kvantovaya Électronika **25**, 586 (1998).
[5] A. V. Kim, M. Yu. Ryabikin, and A. M. Sergeev, Usp. Fiz. Nauk **169**, 58 (1999).
[6] M. V. Fedorov, Usp. Fiz. Nauk **169**, 66 (1999).
[7] V. B. Berestetzkii, E. M. Lifshitz, and L. P. Pitaevskii, *Quantum Electrodynamics*, Pergamon, New York (1984).

Translation provided by the Russian Editorial office

Multiphoton ionization of the hydrogen atom by a circularly polarized electromagnetic field

O. B. Prepelitsa*)

Institute of Applied Physics, Moldavian Academy of Sciences, 277028 Kishinev, Republic of Moldova
(Submitted 28 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 807–820 (September 1999)

This paper examines the multiphoton ionization of the ground state of the hydrogen atom in the field of a circularly polarized intense electromagnetic wave. To describe the states of photoelectrons, quasiclassical wave functions are introduced that partially allow for the effect of an intense electromagnetic wave and that of the Coulomb potential. Expressions are derived for the angular and energy distributions of photoelectrons with energies much lower than the ionization potential of an unperturbed atom. It is found that, due to allowance for the Coulomb potential in the wave function of the final electron states, the transition probability near the ionization threshold tends to a finite value. In addition, the well-known selection rules for multiphoton transitions in a circularly polarized electromagnetic field are derived in a natural way. Finally, the results are compared with those obtained in the Keldysh–Faisal–Reiss approximation. © *1999 American Institute of Physics.* [S1063-7761(99)00509-0]

1. INTRODUCTION

The process of multiphoton ionization of atoms by the field of an intense electromagnetic wave has been studied both theoretically[1–13] and experimentally[14–19] (see also the books cited in Refs. 20–23 and the review articles cited in Refs. 24 and 25). It is only proper to mention the pioneering work of Keldysh[1] and the papers that followed,[2–7] where the main laws governing the multiphoton photoelectric effect were investigated. A common feature of Refs. 1–9 is that the wave function of the photoelectron final state was taken in the form of the wave function of a free electron in the field of the wave (the Volkov wave function), an approach justified only for systems bound by short-range forces. The well-known Keldysh–Faisal–Reiss approximation, which ignores the effect of the Coulomb potential on the final electron state, gives rise to an error in the coefficient of the formula for the probability of ionization of a neutral atom by an intense electromagnetic field and to the wrong threshold dependence of the probability on the photoelectron velocity. The probability of a photoelectron emerging with zero momentum calculated in the Keldysh–Faisal–Reiss approximation proves to be zero.[8] It is known, however, that when a system bound by Coulomb forces breaks up the threshold value of the probability of the process is always finite.[26,27] Moreover, ignoring the Coulomb potential in the case of an atom being ionized by a circularly polarized electromagnetic field makes it impossible to obtain the multiphoton selection rules for the magnetic quantum number. (Because a circularly polarized photon has a projection of angular momentum equal to 1 or -1, the selection rule that an s-photon ionization transition must obey is $s=|m_f-m_i|$, where m_i and m_f are the projections of angular momentum of the initial and final states, respectively.)

The first difficulty was overcome by Perelomov and Popov[4] and Nikishov and Ritus,[7] who found that under certain conditions the Coulomb potential can be taken into account by perturbation-theory techniques. This leads to a correct limiting transition to the formula that describes the probability of ionization of an atom by a constant electric field.[26] Several other methods were proposed to resolve the difficulties in the theory, among which the approach using what is known as the pole approximation,[21,12] and a method of allowing for the Coulomb interaction of the photoelectron with the residual ion by replacing the plane wave in the Volkov solution with a Coulomb wave function of the continuous spectrum with the appropriate asymptotic behavior are worthy of mention.[10,11]

In this paper we will consider the multiphoton ionization of the hydrogen atom by a circularly polarized electromagnetic wave. Quasiclassical ideas will be used to construct a quasistationary wave function of the photoelectron, which will allow for both the intense electromagnetic wave and the Coulomb potential. We will find that in expanding this wave function with appropriate asymptotic properties (an analog of the Sommerfeld function) into a series of partial waves with well-defined values of the projection of angular momentum, transitions are possible only to states that obey multiphoton selection rules for the magnetic quantum number. We will also derive expressions for the angular and energy distributions of the photoelectrons. It will be demonstrated that at the reaction threshold the ionization probability tends to a constant finite value. We will find the critical value of the electromagnetic-field strength at which the height of the main (first) above-barrier peak in the energy distribution of the photoelectrons is equal to that of the second above-barrier peak. The formulas for the ionization probability will be compared with those obtained in the Keldysh–Faisal–Reiss approximation. Finally, we will show that allowing consistently for the Coulomb potential in the wave functions of the final electron state leads to an exponential increase in the ionization probability over that predicted by the Keldysh–Faisal–Reiss theory.

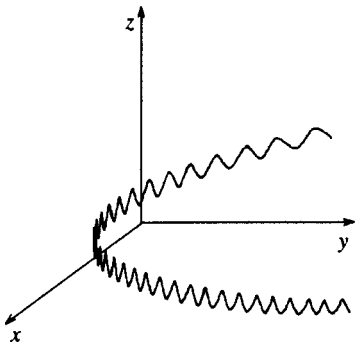

FIG. 1. The path of a classical electron in the field of an intense electromagnetic wave.

2. QUASICLASSICAL WAVE FUNCTIONS OF AN ELECTRON IN THE FIELD OF A STRONG ELECTROMAGNETIC WAVE

We will construct the approximate electron wave functions describing continuous-spectrum states in the Coulomb and electromagnetic fields. Strictly speaking, all states in the field of a wave belong to the continuous spectrum, so that here we are dealing with quasistationary states, and the real part of the average energy of these states is positive. Let us examine the electron states in the near-threshold region with momenta $p \ll \hbar/r_B$, where $r_B=\hbar^2/Me^2$ is the Bohr radius, with M and e the electron mass and charge, respectively. It is well known that in the near-threshold energy region the electron states are quasiclassical. Thus, the main contribution to the electron evolution is provided by a narrow pencil of Feynman paths that lie near the classical path of the electron. Therefore, it is advisable to study more closely the motion of a classical electron in the field of the wave. It is known that if in the absence of an electromagnetic field the electron moves along a certain path $\mathbf{r}(t)$, then in a field $\mathbf{E}(t)$ of frequency ω and under the conditions that

$$\min|\mathbf{r}(t)| \gg \max|\boldsymbol{\alpha}(t)|, \tag{1}$$

$$\omega \gg \frac{1}{T}, \tag{2}$$

with

$$\boldsymbol{\alpha}(t)=\frac{e\mathbf{E}(t)}{M\omega^2} \tag{3}$$

(T is the period of unperturbed motion), the electron moves along the quasistationary path (for details see §30 of Ref. 28):

$$\mathbf{r}'(t)=\mathbf{r}(t)-\boldsymbol{\alpha}(t). \tag{4}$$

We see that the effect of a high-frequency wave on the electrons amounts to oscillations appearing in the motion of the electron along the unperturbed path $\mathbf{r}(t)$ (see Fig. 1). Formally this is equivalent to the electron motion in the absence of an external field but in a noninertial reference frame, with the new and old coordinates related by (4). Hence, basing our reasoning on quasiclassical ideas, we conclude that the wave function of an electron in the field of the wave is approximately the Coulomb wave function written in the noninertial reference frame:

$$\Psi(\mathbf{r},t)=\Psi^C(\mathbf{r}-\boldsymbol{\alpha}(t),t)\exp\left\{-\frac{i}{\hbar}\int_{-\infty}^{t}d\tau\,\frac{e^2\mathbf{A}^2(\tau)}{2Mc^2}\right\}, \tag{5}$$

where $\mathbf{A}(t)$ is the vector potential of the applied electromagnetic field written in the dipole approximation, c is the speed of light in vacuum, and $\Psi^C(\mathbf{r},t)$ is the Coulomb wave function. Note that the function (5) is not a state with a well-defined angular momentum (the same is true of the exact wave function of an electron in the Coulomb and electromagnetic fields). In view of the axial symmetry of the system, the projection of the orbital angular momentum on the symmetry axis is conserved.

The expression (5) can be obtained in a more formal way. We write the electron Hamiltonian in the Coulomb and electromagnetic fields,

$$H=\frac{\left(\hat{\mathbf{P}}-\dfrac{e}{c}\mathbf{A}(t)\right)^2}{2M}-\frac{e^2}{|\mathbf{r}|},$$

and represent it in the following manner:

$$H=H_0+V_{\rm int}, \tag{6}$$

where

$$H_0=\frac{\left(\hat{\mathbf{P}}-\dfrac{e}{c}\mathbf{A}(t)\right)^2}{2M}-\frac{e^2}{|\mathbf{r}-\boldsymbol{\alpha}(t)|}, \tag{7}$$

$$V_{\rm int}=\frac{e^2}{|\mathbf{r}-\boldsymbol{\alpha}(t)|}-\frac{e^2}{|\mathbf{r}|}. \tag{8}$$

Clearly, Eq. (4) is the solution of the classical equations of motion that follow from the Hamiltonian (7), so that it is mainly (7) that determines the classical path of the electron in the field of the wave. In this sense (7) is the zeroth-order Hamiltonian and (8) must be interpreted as a perturbation. Allowing for the fact that the electron is quasiclassical, i.e., that it is almost certain to be near the classical path, we conclude that when go over to the quantum mechanical description the Hamiltonian can still be represented by the sum (6) of the zeroth-order Hamiltonian (7) and the perturbation (8). Hence for the zeroth-order wave functions we can use the solution of the Schrödinger equation with the Hamiltonian (7):

$$i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t}=H_0\Psi(\mathbf{r},t). \tag{9}$$

We seek the solution of Eq. (9) with the initial condition $\Psi(\mathbf{r},-\infty)=\Psi^C(\mathbf{r},-\infty)$ in the form

$$\Psi(\mathbf{r},t)=U(\hat{\mathbf{P}},t)\Psi^C(\mathbf{r},t), \tag{10}$$

where

$$U(\hat{\mathbf{P}},t)=\exp\left\{-\frac{i}{\hbar}\int_{-\infty}^{t}d\tau\left(-\frac{e\mathbf{A}(\tau)\cdot\hat{\mathbf{P}}}{Mc}+\frac{e^2\mathbf{A}^2(\tau)}{2Mc^2}\right)\right\}. \tag{11}$$

Substituting (10) and (11) in Eq. (9) and performing simple transformations, we arrive at an equation for $\Psi^C(\mathbf{r},t)$:

$$i\hbar\frac{\partial\Psi^C(\mathbf{r},t)}{\partial t}=\left(\frac{\hat{\mathbf{P}}^2}{2M}-\frac{e^2}{|\mathbf{r}|}\right)\Psi^C(\mathbf{r},t).$$

This implies that $\Psi^C(\mathbf{r},t)$ is indeed a Coulomb wave function that determines the electron state in the absence of an external electromagnetic field. If we take formula (3) into account, we can easily show that the expressions (5) and (10) are identical.

3. PROBABILITY OF MULTIPHOTON IONIZATION OF THE GROUND STATE OF THE HYDROGEN ATOM

Let us study the probability amplitude of a transition from the ground state of the unperturbed atom, $\Psi_i^C(\mathbf{r},t)$, to the continuous-spectrum states $\Psi_f(\mathbf{r},t)$. The expression for this amplitude is

$$A_{if}=-\frac{i}{\hbar}\int_{-\infty}^{\infty}dt\int d\mathbf{r}\Psi_f^*(\mathbf{r},t)\times\left(-\frac{e\mathbf{A}(t)\cdot\hat{\mathbf{P}}}{Mc}+\frac{e^2\mathbf{A}^2(t)}{2Mc^2}\right)\Psi_i^C(\mathbf{r},t). \tag{12}$$

According to (5), (10), and (11), the final electron state can be expressed in terms of the unperturbed Coulomb wave function of the continuous spectrum, $\Psi_f^C(\mathbf{r},t)$, in two equivalent ways:

$$\Psi_f(\mathbf{r},t)=\Psi_f^C(\mathbf{r}-\boldsymbol{\alpha}(t))\times\exp\left\{-\frac{i}{\hbar}\left(\frac{p^2}{2M}t+\int_{-\infty}^{t}d\tau\frac{e^2\mathbf{A}^2(\tau)}{2Mc^2}\right)\right\}, \tag{13}$$

$$\Psi_f(\mathbf{r},t)=U(\hat{\mathbf{P}},t)\Psi_f^C(\mathbf{r})\exp\left\{-\frac{i}{\hbar}\frac{p^2}{2M}t\right\}. \tag{14}$$

We will assume that at times $t\to\pm\infty$ the field is zero. Then, using the representation (14) and doing the necessary transformations related to the integration by parts in (12), we obtain

$$A_{if}=\frac{i}{\hbar}\left(\frac{p^2}{2M}+I_0\right)\int_{-\infty}^{\infty}dt\int d\mathbf{r}\Psi_f^*(\mathbf{r},t)\Psi_i^C(\mathbf{r},t),$$

where I_0 is the ionization potential of the ground state of the atom.

We will now introduce the Fourier transforms of the Coulomb wave functions:

$$\varphi_{f,i}(\mathbf{k})=\frac{1}{(2\pi)^3}\int d\mathbf{r}\Psi_{f,i}^C(\mathbf{r})\exp\{-i\mathbf{k}\cdot\mathbf{r}\}.$$

Using these and the representation (13) of the final-state wave function, we can transform the ionization amplitude into

$$A_{if}=\frac{i}{\hbar}\left(\frac{p^2}{2M}+I_0\right)\int_{-\infty}^{\infty}dt\exp\left\{\frac{i}{\hbar}\left[\left(\frac{p^2}{2M}+I_0\right)t+\int_{-\infty}^{t}d\tau\frac{e^2\mathbf{A}^2(\tau)}{2Mc^2}\right]\right\}\Phi(\boldsymbol{\alpha}(t)), \tag{15}$$

$$\Phi(\boldsymbol{\alpha}(t))=\int d\mathbf{r}\,d\mathbf{k}\,d\mathbf{k}'\varphi_f^*(\mathbf{k})\varphi_i(\mathbf{k}')\times\exp\{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}\}\exp\{-i\mathbf{k}\cdot\boldsymbol{\alpha}(t)\}. \tag{16}$$

Clearly, integration over the coordinate space yields a delta function $\delta(\mathbf{k}-\mathbf{k}')$, and this resolves all problems with integration over $\mathbf{k}'$. Formula (16) becomes

$$\Phi(\boldsymbol{\alpha}(t))=(2\pi)^3\int d\mathbf{k}\varphi_f^*(\mathbf{k})\varphi_i(\mathbf{k})\exp\{-i\mathbf{k}\cdot\boldsymbol{\alpha}(t)\}. \tag{17}$$

We can use the following trick to evaluate (17). Instead of the specified bound–free transition, we examine a bound–bound transition from the ground state to a state belonging to the discrete spectrum. For the time being we interpret the $\varphi_j(\mathbf{k})$ in (17) as the Fourier transform of the wave function of a bound state with quantum numbers (nlm). After evaluating the integral by analytically continuing n into the domain of imaginary values, we can find (17). Such a procedure simplifies the calculations significantly, since in evaluating (17) it is much simpler to deal with Fourier transforms of bound states than with Fourier transforms of the continuous-spectrum states. Note that a similar trick was used in Refs. 29 and 30, where the amplitude of a bound–free transition was found by analytic continuation that involved a known bound–bound transition amplitude.

The wave function of the bound state into which the system is formally transferred can be written

$$\overline{\Psi}_f^C(\mathbf{r})=R_{nl}(r)Y_{lm}(\theta,\varphi). \tag{18}$$

Next we write the Fourier transforms of the ground state and state (18) (see Ref. 31):

$$\varphi_i(\mathbf{k})=\frac{1}{\sqrt{\pi r_B^5}}\frac{1}{(k^2+1/r_B^2)^2}, \tag{19}$$

$$\overline{\varphi}_f(\mathbf{k})=F_{nl}(k)Y_{lm}(\theta,\varphi), \tag{20}$$

$$F_{nl}(k)=\frac{2^{2l+1}l!k^l}{\pi^2r_B^{l+5/2}n^{l+2}(k^2+1/(nr_B)^2)^{l+2}}\times C_{n-l-1}^{l+1}\left(\frac{k^2-1/(nr_B)^2}{k^2+1/(nr_B)^2}\right), \tag{21}$$

where $C_N^n(x)$ is the Gegenbauer function. We will assume that the applied electromagnetic field is circularly polarized:

$$\mathbf{E}(t)=E_0(\mathbf{e}_x\cos\omega t+\mathbf{e}_y\sin\omega t), \tag{22}$$

where $\mathbf{e}_x$ and $\mathbf{e}_y$ are the unit vectors of a Cartesian system of coordinates. We will also assume that the atom is oriented in such a way that the quantization axis coincides with the direction in which the electromagnetic wave propagates. Thus, taking Eqs. (17) and (19)–(22) into account, we can write

$$\overline{\Phi}(\boldsymbol{\alpha}(t))=\frac{(2\pi)^3}{\sqrt{\pi r_B^5}}\int_0^\infty dk\,k^2\frac{F_{nl}(k)}{(k^2+1/r_B^2)^2}\int d\Omega Y_{lm}^*(\theta,\varphi)$$
$$\times\exp\{-ik\alpha_0\sin\theta\cos(\varphi-\omega t)\},$$

where $\alpha_0=eE_0/M\omega^2$. It can be demonstrated that the value of the integral is largely determined by the poles of the integrand. According to (21), the poles are at $k=i/r_B$ and $k=i/nr_B$. In weak fields, where we have $\alpha_0\ll r_\beta$, the contributions of the two poles are comparable, while in strong fields, $\alpha_0\gg r_B$, the main contribution is provided by the pole that lies closes to the real axis, $k=i/nr_B$ (due to the exponential factor in the integrand). In the latter case we ignore the small terms of order $1/n$, with $n\gg1$, and take the denominator $(k^2+1/r_B^2)^2$ outside the integral sign at the point $k=i/nr_B$. The result is

$$\overline{\Phi}(\boldsymbol{\alpha}(t))=(2\pi)^3\varphi_i\left(\frac{i}{nr_B}\right)\int_0^\infty dk\,k^2F_{nl}(k)$$
$$\times\int d\Omega Y_{lm}^*(\theta,\varphi)$$
$$\times\exp\{-ik\alpha_0\sin\theta\cos(\varphi-\omega t)\}.$$

For the sake of comparison we will write the inverse Fourier transform of the wave function (18):

$$\overline{\Psi}_f^C(\mathbf{r}')=\int_0^\infty dk\,k^2F_{nl}(k)\int d\Omega Y_{lm}(\theta,\varphi)$$
$$\times\exp\{ikr'[\cos\theta\cos\theta'$$
$$+\sin\theta\sin\theta'\cos(\varphi-\varphi')]\}.$$

Clearly, the penultimate expression is in fact the inverse Fourier transform of the function $\overline{\Psi}_f^{C*}(\mathbf{r}')$ specified at the point with $r'=\alpha_0$, $\theta'=\pi/2$, and $\varphi'=\omega t$. Thus, allowing for (18), we have

$$\overline{\Phi}(\boldsymbol{\alpha}(t))=(2\pi)^3\varphi_i\left(\frac{i}{nr_B}\right)R_{nl}(\alpha_0)Y_{lm}\left(\frac{\pi}{2},\omega t\right),$$

$$\frac{\alpha_0}{r_B},\quad n\gg1.$$

If we continue this expression analytically into the domain of imaginary values of n, we arrive at the case where ionization proceeds to a state for which the angular momentum and its projection are well-defined. (More precisely, $\Psi_f^C(\mathbf{r},t)$ in (13) and (14) is a state with a well-defined angular momentum and its projection.) But according to the general principles of scattering theory, for the Coulomb wave function in (13) and (14) we must take a wave function whose asymptotic form contains a plane wave and a converging spherical wave. This Coulomb wave function can be written in the form of a partial-wave expansion:[26]

$$\Psi_f^C(\mathbf{r})=\frac{1}{\sqrt{2\pi\hbar}p}\sum_{l=0}^\infty\sum_{m=-l}^{l}i^l\exp\{-i\delta_l\}R_{pl}(r)$$
$$\times Y_{lm}(\theta,\varphi)Y_{lm}^*(\theta_p,\varphi_p),$$

where

$$R_{pl}(r)=\frac{C_{pl}}{(2l+1)!r_B}\left(\frac{2pr}{\hbar}\right)^l\exp\left\{-\frac{i}{\hbar}pr\right\}$$
$$\times F\left(l+1+i\vartheta,2l+2,2\frac{i}{\hbar}pr\right),$$

$$C_{pl}=\sqrt{\frac{8\pi}{\vartheta(1-\exp\{-2\pi\vartheta\})}}\prod_{\eta=1}^{l}\sqrt{\eta^2+\vartheta^2},$$

$$\vartheta=\frac{\hbar}{pr_B},$$

with θ_p and φ_p the axial and polar angles of vector $\mathbf{p}$, respectively.

Thus, allowing for the above formula, we arrive at the desired value of the integral (17):

$$\Phi(\boldsymbol{\alpha}(t))=\frac{(2\pi)^{5/2}}{\hbar^{1/2}p}\varphi_i\left(\frac{i}{nr_B}\right)$$
$$\times\sum_{l=0}^\infty\sum_{m=-l}^{l}i^{-l}\exp\{i\delta_l\}R_{pl}(\alpha_0)$$
$$\times Y_{lm}\left(\frac{\pi}{2},\omega t\right)Y_{lm}^*(\theta_p,\varphi_p).$$

Using the relationship that exists between spherical functions and associated Legendre polynomials,

$$Y_{lm}(\theta,\varphi)=(-1)^{(m+|m|)/2}i^l\sqrt{\frac{2l+1}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}}$$
$$\times P_l^{|m|}(\cos\theta)\exp\{im\varphi\},$$

and formula (15) and performing a simple integration, we find the ionization probability amplitude:

$$A_{if}=i\frac{(2\pi)^{7/2}}{\hbar^{1/2}p}\left(\frac{p^2}{2M}+I_0\right)\varphi_i\left(\frac{i}{nr_B}\right)$$
$$\times\sum_{l=0}^\infty\sum_{m=-l}^{l}\frac{2l+1}{4\pi}\exp\left\{i\left(\delta_l+\frac{\pi}{2}(l+m+|m|)\right)\right\}$$
$$\times\frac{(l-|m|)!}{(l+|m|)!}P_l^{|m|}(0)P_l^{|m|}(\cos\theta_p)\exp\{-im\varphi_p\}R_{pl}(\alpha_0)$$
$$\times\delta\left(\frac{p^2}{2M}+I_0+\frac{e^2E_0^2}{2M\omega^2}-m\hbar\omega\right),\qquad(23)$$

where it is assumed that α_0/r_B and $\hbar/pr_B$ are much larger than unity.

Let us examine the above expression more thoroughly. The presence of a delta function means that of all the partial waves only that in which the magnetic quantum number coincides with the number of photons involved in the process can serve as the final state. This agrees with the selection rule for magnetic quantum numbers in dipole multiphoton transitions (from the ground state) initiated by circularly polarized radiation.[20]

Allowing for the property of the zeros of Legendre polynomials, $P_l^{|m|}(0)\sim\cos[(l+|m|)\pi/2]$, we see that the transition can end only in states in which $l+|m|=2N$, N

$=1,2,3...$. On the other hand, the dipole selection rules for the orbital quantum number require that its variation be equal to the number of circularly polarized photons absorbed in the transition process, i.e., $l=|m|$. Nevertheless, there is no contradiction here, since the selection rules in question have been established for states with well-defined angular momenta. In our case, because of the perturbation of the final states by the electromagnetic wave, the number l is not a constant of motion. Hence the terms in (23) with given l do not determine the probability amplitude of ionization to a state with a well-defined angular momentum equal to $l\hbar$. This explains why the selection rules break down in (23).

We will assume that

$$\frac{p\alpha_0}{l\hbar}\ll 1 \tag{24}$$

[we will shortly show that this inequality follows from the condition for the applicability of model (1)]. For the functions $R_{pl}(\alpha_0)$ in (23) we employ the asymptotic representation

$$R_{pl}(\alpha_0)\approx \frac{C_{pl}}{r_B(2l+1)!}\left(\frac{2p\alpha_0}{\hbar}\right)^l, \quad \frac{p\alpha_0}{l\hbar}\ll 1. \tag{25}$$

With this in mind we can now show that the main contribution to the ionization amplitude is provided by the term with the smallest possible value of l:

$$l=|m|=s, \tag{26}$$

where by s we have denoted the number of photons absorbed by the electron in the transition. Thus, discarding the unimportant phase factors, we can write the s-photon ionization probability amplitude:

$$A_{if}=2\pi\sum_{s=[s_0]} D_{ps}\left(\frac{p^2}{2M}+I_0\right)\left(\frac{p\alpha_0}{\hbar}\sin\theta\right)^s \times\delta\left(\frac{p^2}{2M}+I_0+\frac{e^2E_0^2}{2M\omega_0^2}-s\hbar\omega\right),$$

$$D_{ps}=\sqrt{\frac{\pi^5 r_B^3\vartheta}{\hbar^3}}\,\frac{1}{2^{s-2}s!}\prod_{\eta=1}^{s}\sqrt{1+\left(\frac{\vartheta}{\eta}\right)^2},$$

where

$$s_0=\frac{I_0}{\hbar\omega}+\frac{e^2E_0^2}{2M\hbar\omega^3},$$

$[s_0]$ is the minimum number of the electromagnetic-field photons needed to ionize the atom, and θ is the angle between the photoelectron momentum and the direction of wave propagation. In deriving these formulas we used Eq. (19) and allowed for the fact that the photoelectron momenta are small compared to the Bohr momentum: $pr_B/\hbar\ll 1$.

For the finite product we have the asymptotic representation

$$\prod_{\eta=1}^{s}\left(1+\left(\frac{\vartheta}{\eta}\right)^2\right)=\frac{1}{2\pi\vartheta}\left(1+\left(\frac{\vartheta}{s}\right)^2\right)^{s+1/2} \times\exp\left\{-2\vartheta\arctan\frac{\vartheta}{s}+\pi\vartheta\right\}$$

with $\vartheta,s\gg 1$. This expression simplifies significantly when $\vartheta/s\ll 1$:

$$\prod_{\eta=1}^{s}\left(1+\left(\frac{\vartheta}{\eta}\right)^2\right)=\frac{1}{2\pi\vartheta}\exp\{\pi\vartheta\}. \tag{27}$$

Now, performing well-known transformations, we find an expression for the differential multiphoton ionization probability per unit time (the ionization rate) in the form of a sum of partial probabilities:

$$\frac{dW}{d\Omega}=\sum_{s=[s_0]}\frac{dW^s}{d\Omega}, \tag{28}$$

$$\frac{dW^s}{d\Omega}=\frac{(2\pi)^6}{2^{2(s+1)}(s!)^2}\,\frac{\hbar\omega^2}{I_0}\left(s-\frac{e^2E_0^2}{2M\hbar\omega^3}\right)^2 \times\left(\frac{p_s\alpha_0}{\hbar}\sin\theta\right)^{2s}g(p_s), \tag{29}$$

where

$$g(p_s)=\prod_{\eta=1}^{s}\left(1+\frac{\hbar^2}{p_s^2r_B^2\eta^2}\right), \quad p_s=\sqrt{2M\hbar\omega(s-s_0)}.$$

Here and in what follows we assume that $1\ll\alpha_0/r_B\ll s$ and $p_s\alpha_0/\hbar\ll s$.

Equations (27) and (28) imply that the angular distribution of the photoelectrons has a narrow peak lying in the polarization plane of the electromagnetic wave, $\theta=\pi/2$. As we move away from the angle $\theta=\pi/2$, the probability of detecting a photoelectron decreases rapidly and vanishes in the limit $\theta=0$. Thus, qualitatively the angular distribution of photoelectrons follows the same pattern as it does in the Keldysh–Faisal–Reiss approximation.[8]

Integrating the differential probability over the solid angle, we find the photoelectron energy distribution:

$$W=\sum_{s=[s_0]}W^s, \tag{30}$$

$$W^s=\frac{(2\pi)^7}{2(2s+1)!}\,\frac{\hbar\omega^2}{I_0}\left(s-\frac{e^2E_0^2}{2M\hbar\omega^3}\right)^2\left(\frac{p_s\alpha_0}{\hbar}\right)^{2s}g(p_s). \tag{31}$$

Formula (31) shows that near the ionization threshold, where the photoelectron momentum tends to zero, the probability of the process remains finite, as it should in reactions with charged-particle production:[26,27]

$$\lim_{p_{s_0}\to 0}W^{[s_0]}=\frac{(2\pi)^7}{2(2[s_0]+1)!}\,\frac{I_0}{\hbar}\left(\frac{\alpha_0}{r_B}\right)^{2[s_0]}. \tag{32}$$

Let us find the critical value of the electromagnetic-field strength at which the height $s=[s_0]+1$ of the first additional peak becomes equal to the height of the main peak $s=[s_0]$. We will assume that the height of the main peak is described by (32). Then, using (31) and (27), we arrive at the ratio

$$\frac{W^{[s_0+1]}}{W^{[s_0]}}\approx\frac{1}{\pi}\left(\frac{\hbar\omega}{I_0}\right)^{[s_0]-1/2}\frac{e^2E_0^2}{M\hbar\omega^3}\,\exp\left\{\pi\sqrt{\frac{I_0}{\hbar\omega}}\right\}.$$

Now, from the condition that $W^{[s_0+1]}/W^{[s_0]}=1$ we can easily derive an expression for the critical value of the electromagnetic field strength:

$$E_0^{\rm cr}\approx\sqrt{\frac{\pi\hbar\omega^3}{e^2}}\left(\frac{I_0}{\hbar\omega}\right)^{[s_0]/2-1/4}\exp\left\{-\frac{\pi}{2}\sqrt{\frac{I_0}{\hbar\omega}}\right\}.$$

For instance, for an ionization potential $I_0=13.6$ eV and a field frequency $\omega\approx2.92\times10^{15}\,{\rm s}^{-1}$ we have $E_0^{\rm cr}\approx9.76\times10^7$ V/cm. For large photoelectron threshold momenta, $p_{s_0}^2/2M\sim\hbar\omega$, the critical field value is of the same order as in the Keldysh–Faisal–Reiss theory. For $E_0>E_0^{\rm cr}$ the photoelectron energy distribution has many peaks, and the distribution maximum is shifted to the right from the main peak $s=[s_0]$. Note that far from the ionization threshold, $s-[s_0]\gg1$, the probability ratio has a more usual form:

$$\frac{W^{s+1}}{W^s}\sim\frac{E_0^2}{\omega^2},$$

which differs from the quasiclassical parameter $E_0^2/\omega^{10/3}$. This discrepancy arises because in finding the wave functions (5) and (10) we ignored the interaction (8).

Let us establish the number of photons $s_{\max}$ at which for a given field strength E_0 the ionization probability (31) reaches its maximum.

Using the well-known Stirling formula and Eqs. (31) and (27), we can write

$$W^s=\frac{(2\pi)^5}{2(2s+1)}\sqrt{\frac{\pi(s-s_0)}{s}}\,\frac{\hbar^2\omega^3}{I_0^2}\times\left(s-\frac{e^2E_0^2}{2M\hbar\omega^3}\right)^2\exp\{f(s)\},$$

$$f(s)=2s\left(\ln\frac{p_s\alpha_0}{2s\hbar}+1\right)+\pi\sqrt{\frac{I_0}{\hbar\omega(s-s_0)}},\quad s>[s_0].$$

If we now (formally) assume that s is a continuous parameter, we can find $s_{\max}$ by solving the equation $f'(s_{\max})=0$. In explicit form this equation is

$$\ln\frac{e^2E_0^2}{2M\hbar\omega^3}=\ln\frac{s_{\max}^2}{s_{\max}-s_0}-\frac{1}{s_{\max}-s_0}\times\left(s_{\max}-\frac{\pi}{2}\sqrt{\frac{I_0}{\hbar\omega(s_{\max}-s_0)}}\right).\qquad(33)$$

We see that this formula determines (implicitly) the function $s_{\max}=s_{\max}(E_0)$. Unfortunately, there is no way to solve Eq. (33) can be solved. For the sake of an example we give the value of $s_{\max}$ for the following values of the parameters: $I_0\approx13.6$ eV, $\omega\approx2\times10^{15}\,{\rm s}^{-1}$, $[s_0]=22$, and $E_0\approx1.5\times10^8$ V/cm. In accordance with (33) we have $s_{\max}=25$.

The maximum value of the probability, $W^{s_{\max}}$, is given by the expression

$$W^{s_{\max}}=\frac{(2\pi)^5}{2(2s_{\max}+1)}\sqrt{\frac{\pi(s_{\max}-s_0)}{s_{\max}}}\,\frac{\hbar^2\omega^3}{I_0^2}\times\left(s_{\max}-\frac{e^2E_0^2}{2M\hbar\omega^3}\right)\exp\{f(s_{\max})\},\qquad(34)$$

$$f(s_{\max})=-\frac{s_{\max}}{s_{\max}-s_0}\left(2s_0-s_{\max}-\frac{\pi}{2}\sqrt{\frac{I_0}{\hbar\omega(s_{\max}-s_0)}}\times\left(1+2\frac{s_{\max}-s_0}{s_{\max}}\right)\right),$$

$[s_0]\gg s_{\max}-[s_0]>1$.

The two formulas, (33) and (34) can be used to compare the theoretical and experimental dependences of $W^{s_{\max}}$ on the field strength.

It is interesting to compare the s-photonionization probability (31) with the probability of a similar process considered in the Keldysh–Faisal–Reiss approximation:[8]

$$W_{\rm KFR}^s=\frac{2^5\pi^4}{(2s+1)!}\sqrt{\frac{\hbar^3\omega^5}{I_0^3}}\left(s-\frac{e^2E_0^2}{2M\hbar\omega^3}\right)^2\times\sqrt{s-\frac{I_0}{\hbar\omega}-\frac{e^2E_0^2}{2M\hbar\omega^3}}\left(\frac{p_s\alpha_0}{\hbar}\right)^{2s}.\qquad(35)$$

in the limit $\hbar/p_sr_Bs\ll1$ we have

$$\frac{W^s}{W_{\rm KFR}^s}=\exp\left\{\pi\sqrt{\frac{I_0}{\hbar\omega(s-s_0)}}\right\},\quad s>[s_0].$$

Thus, even partial allowance for the Coulomb potential in the final electron state leads to a sharp (exponential) increase in the transition probability. The exponential differences of the formulas (31) and (35) are due to the long range of Coulomb forces, which were ignored in (35). Note that when s is large, more precisely, in the limit $s\gg s_0+\pi^2I_0/\hbar\omega$, the probability ratio $W^{(s)}/W_{\rm KFR}^{(s)}$ tends to unity. This was to be expected, since for fast photoelectrons, whose energy is much higher than the photoionization potential, the Born approximation is valid. This means that the effect of the Coulomb potential on the wave functions of the photoelectrons can be ignored, which in turn means that we can use the Keldysh–Faisal–Reiss approximation.

The reader will recall that the main condition for the applicability of Eqs. (28)–(34) is the quasiclassicality of the final electron state, i.e., the smallness of the electron momentum in comparison to the Bohr momentum. The results we have arrive at give a good description of the case of multiphoton ionization of the atom, $\gamma\gg1$, where the Keldysh parameter is given by the formula $\gamma^2=2MI_0\omega^2/e^2E_0^2$, since here the probability of a photoelectron having a momentum of order of, or larger than, the Bohr momentum is negligible. In the adiabatic case $\gamma\ll1$, the important photoelectron momenta (near the maximum of the energy distribution) are much larger than the Bohr momentum.[25,13] Hence for $\gamma\ll1$, the formulas (28)–(34) can be used only to describe the low-energy fraction of the photoelectrons. Unfortunately, this makes it impossible to calculate the total probability of ionization in ultrahigh fields and thus to pass to the limiting case of ionization of an atom by a constant electric field.

In conclusion we will examine the criteria of applicability of (1) and (2) in greater detail. Using the formula that gives the minimum distance between the nucleus and the classical electron in the Coulomb problem,[28]

$$r_{\min}=\frac{\hbar^2}{p^2 r_B}\left(\sqrt{1+\frac{p^2 r_B^2}{\hbar^2}l^2}-1\right), \quad l\gg 1,$$

we can write the inequality (1) in the form

$$\frac{\alpha_0}{\sqrt{1+p^2 r_B^2 l^2/\hbar^2}-1}\,\frac{p^2 r_B}{\hbar^2}\ll 1. \tag{36}$$

In the limit $pr_B l/\hbar\gg 1$ this reduces to the $p\alpha_0/\hbar l\ll 1$. Clearly, this inequality determines the restriction on the electromagnetic-field strength that follows from (36) and coincides with condition (24), which governs the main results (28)–(34).

The motion of an electron with positive energy in the field of the nucleus is not periodic, with the result that the criterion (2) must be reformulated. In our case, the criterion (2) corresponds to the condition that the electron displacement along an unperturbed path in the course of the period of the electromagnetic field be much smaller than the distance to the nucleus:

$$\frac{2\pi}{\omega}\,\frac{p}{M}\ll r_{\min}$$

or, to put it differently,

$$\frac{2\pi}{\omega}\,\frac{p^3 r_B}{\hbar^2 M}\,\frac{1}{\sqrt{1+p^2 r_B^2 l^2/\hbar^2}-1}\ll 1. \tag{37}$$

It can be demonstrated that for $l\gg 1$ and $pr_B/\hbar\ll 1$ the conditions (36) and (37) can be met simultaneously over a wide range of values of the parameters ω and E_0.

4. CONCLUSION

We have studied the above-threshold ionization of an atom by a circularly polarized electromagnetic field and found that the use of quasiclassical wave functions (5) and (10) of an electron in the field of the wave has certain merits in comparison to the use of the Volkov wave function or the functions used by Shakeshaft and Potvliege[10] and Basile *et al.*[11] By allowing the Coulomb interaction in (5) and (10), we were able to arrive at two important results: first, the finite threshold value of the ionization probability (32) and, second, the appropriate selection rules (26) for the magnetic quantum number. As noted earlier the Keldysh–Faisal–Reiss approximation does not lead to such results. The violation of the first condition in this approximation is obvious [see Eq. (35)], while the violation of the selection rules can be made clearer if in the expression for the transition amplitude

$$A_{if}\sim\int_{-\infty}^{\infty}dt\int d\mathbf{r}\exp\left\{-\frac{i}{\hbar}\left[\mathbf{p}\cdot\mathbf{r}-\frac{1}{2M}\times\int_{-\infty}^{t}d\tau\left(\mathbf{p}-\frac{e}{c}\mathbf{A}(\tau)\right)^2\right]\right\}\exp\left\{-\frac{r}{r_B}\right\}$$

we expand the plane wave in a series in spherical functions[26] and in evaluating the integral allow for the orthogonality of these functions.

*)E-mail: exciton@cc.acad.md

[1] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1965)].
[2] A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **50**, 1393 (1966) [Sov. Phys. JETP **23**, 924 (1966)].
[3] A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **51**, 309 (1966) [Sov. Phys. JETP **24**, 207 (1967)].
[4] A. M. Perelomov and V. S. Popov, Zh. Éksp. Teor. Fiz. **52**, 514 (1967) [Sov. Phys. JETP **25**, 336 (1967)].
[5] V. S. Popov, V. P. Kuznetsov, and A. M. Perelomov, Zh. Éksp. Teor. Fiz. **53**, 331 (1967) [Sov. Phys. JETP **26**, 222 (1968)].
[6] L. P. Kotova and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **52**, 732 (1967) [Sov. Phys. JETP **25**, 481 (1967)].
[7] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **52**, 223 (1967) [Sov. Phys. JETP **25**, 145 (1967)].
[8] H. R. Reiss, Phys. Rev. A **22**, 1786 (1980).
[9] M. V. Ammosov, N. B. Delone, and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz. **91**, 2008 (1986) [Sov. Phys. JETP **64**, 1191 (1986)].
[10] R. Shakeshaft and R. M. Potvliege, Phys. Rev. A **36**, 5478 (1987).
[11] S. Basile, F. Trombetta, G. Ferrante, R. Burlon and C. Leone, Phys. Rev. A **37**, 1050 (1988).
[12] D. F. Zaretskiĭ and É. A. Nersesov, Zh. Éksp. Teor. Fiz. **103**, 1191 (1993) [JETP **76**, 583 (1993)].
[13] V. P. Kraĭnov and B. Shokri, Zh. Éksp. Teor. Fiz. **107**, 1180 (1995) [JETP **80**, 657 (1995)].
[14] G. S. Voronov, G. A. Delone, and N. B. Delone, Zh. Éksp. Teor. Fiz. **51**, 1660 (1966) [Sov. Phys. JETP **24**, 1122 (1967)].
[15] G. A. Delone, and N. B. Delone, JETP Lett. **10**, 265 (1969).
[16] P. Agostini, G. Barjot, G. Mainfray, and J. Thebault, IEEE J. Quantum Electron. **QE-6**, 12 (1970).
[17] P. Agostini, F. Fabre, G. Mainfray, G. Petite, and N. K. Rahman, Phys. Rev. Lett. **42**, 1127 (1979).
[18] G. Petite, P. Agostini, and F. Yergeau, J. Opt. Soc. Am. B **4**, 765 (1987).
[19] H. Humpert, H. Schurier, R. Hippler, and H. O. Lutz, Phys. Rev. A **32**, 3787 (1985).
[20] N. B. Delone and V. P. Kraĭnov, *Atoms in Strong Light Fields*, Springer-Verlag, Berlin (1985).
[21] M. V. Fedorov, *Interaction of Intense Laser Light with Free Electrons*, Harwood Academic Publishers, Chur (1991).
[22] L. P. Rapoport, B. A. Zon, and N. L. Manakov, *The Theory of Multiphoton Processes in Atoms* [in Russian], Atomizdat, Moscow (1978).
[23] *Multiphoton Ionization of Atoms* [in Russian], N. G. Basov (Ed.), Proc. (Trudy) of the P. N. Lebedev Physics Institute, Vol. 115, Moscow (1980).
[24] N. B. Delone, Usp. Fiz. Nauk **115**, 361 (1975) [Sov. Phys. Usp. **18**, 169 (1975)].
[25] N. B. Delone and M. V. Fedorov, Usp. Fiz. Nauk **158**, 215 (1989) [Sov. Phys. Usp. **32**, 500 (1989)].
[26] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[27] A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions, and Decay in Non-relativistic Quantum Mechanics*, Nauka, Moscow (1971) [English translation of the 1st Russian edition: NASA Techn. Transl. F-510 (1969)].
[28] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., Pergamon Press, Oxford (1976).
[29] L. A. Bureeva, Astron. Zh. **45**, 1215 (1968) [Sov. Astron. **12**, 962 (1969)].
[30] S. P. Goreslavskiĭ, N. B. Delone, and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz. **82**, 1789 (1982) [Sov. Phys. JETP **55**, 1032 (1982)].
[31] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms*, Plenum Press, New York (1977).

Translated by Eugene Yankovsky

Statistical properties of resonance radiation selectively reflected from a cold semi-infinite medium

B. A. Veklenko*) and Yu. B. Sherkunov

Moscow Power-Engineering Institute, 111250 Moscow, Russia
(Submitted 15 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 821–837 (September 1999)

The problem of the quantum-statistical properties of resonance radiation selectively reflected from unexcited media in the case where the photons are mutually correlated in the incident radiation flux is posed and solved. Allowance for mutual photon correlation precludes, in principle, solving the problem by perturbation methods. A quantum analog of the optical theorem of absorption, which causes an infinite subsequence of Feynman diagrams to vanish, is pointed out. The frequency-angle distribution of the photons in the reflected flux is predicted. The Fresnel formulas are reconstructed for the averaged reflection characteristics. The limits for their applicability in describing the reflection of mutually correlated photons are given. A suppression effect is predicted for the reflection of radiation from a laser source. [S1063-7761(99)00609-5]

1. INTRODUCTION

The resonant interaction of an electromagnetic field with matter is classified as a strong interaction in the sense that the optical effects caused by it preclude investigations by perturbative methods. The interest in such phenomena has traditionally been great and has become stronger in recent years as a result of the research on optical effects in dense media[1,2] and on the interaction of laser radiation with matter.[3–5] One of the manifestations of the resonant interaction of an electromagnetic field with matter is the selective reflection of light from an interface between a gas and free space. This area of research has its own rich past. In 1909 Wood[6] experimentally discovered a change from diffuse to specular light scattering as the pressure of the reflecting vapor was increased, providing food for thought to researchers for a good half a century.[7,8] In 1966 Koester[9] experimentally discovered the possibility of amplifying resonance radiation when it is reflected from a medium with population inversion. That work produced a new flood of research, which was summarized to some extent in a monograph by Boĭko and Petrov.[10] It was still too early to speak about complete agreement between theory and experiment. However, just then a thorough investigation of the influence of stimulated emission processes on the reflection mechanism demonstrated[11,12] the inapplicability of the semiclassical theory of radiation based on an unquantized electromagnetic field. It turned out that consideration of the quantum properties of radiation qualitatively alters the calculation results on a macroscopic level. This is ultimately caused by the correlation properties of photons in the medium and the change in the number of photons as a result of the stimulated emission. It is natural to expect that something similar should occur in the reverse process, i.e., in the scattering of correlated photons under conditions allowing their absorption. This means that correlated photons are not scattered independently in cold media, even if the interaction of light with each individual atom of the medium is considered in the linear approximation. In analogy to the effects described in Ref. 12, the inapplicability of perturbation theory should be expected in this case. Why was such a phenomenon not discovered before, what are its characteristic features, and how do the Fresnel formulas, which have no nonanalytic features with respect to charge, arise? Answers to these questions are given below.

We shall examine the resonant reflection of two transversely polarized photons which belong to a mode being scattered with the wave vector $\mathbf{k}_0$ and the polarization λ_0 and can be described by the common wave function $(\hat{\alpha}^+_{\mathbf{k}_0\lambda_0})^2|0\rangle/\sqrt{2}$ from a semi-infinite medium. We shall use $\hat{\alpha}^+_{\mathbf{k}\lambda}$ $(\hat{\alpha}_{\mathbf{k}\lambda})$ to denote the creation (annihilation) operator of a photon in the state $(\mathbf{k},\lambda)$. We shall determine the probability of finding one of both photons in the reflection mode $(\mathbf{k},\lambda)$. We shall study the statistical properties of the reflected field in this elementary example, and we shall investigate the averaged characteristics of the field.

2. PRELIMINARY ANALYSIS

Some qualitative features of the reflection of correlated photons can be discovered on the basis of general arguments. We use φ_i to denote the eigenfunctions of the Hamiltonian of the scattering medium. We take into account the adiabatic hypothesis. Let the medium be in the state φ_0 before the interaction with radiation. The complete wave function of the ''radiation+medium'' system can be represented in the form

$$\Psi = \varphi_0 f_0 + \sum_{i\neq 0} \varphi_i f_i .$$

To describe the radiation field in the medium it is sufficient to know its density matrix:

$$\rho = \mathrm{Tr}_a \Psi\Psi^* = \rho^{(c)} + \rho^{(n)},$$

$$\rho^{(c)} = f_0 f_0^*, \quad \rho^{(n)} = \sum_{i\neq 0} f_i f_i^* . \tag{1}$$

The operation Tr_a extends to the atomic system, and the functions f_i depend on the arguments of the electromagnetic field. The matrix $\rho^{(c)}$, which corresponds to the coherent scattering channel, describes the scattering processes which result in the return of medium atoms to the initial (including the translational) state. The matrix $\rho^{(n)}$, which corresponds to the incoherent scattering channel, describes the other scattering processes. The matrix ρ can be measured directly by photon-counting methods.[13] It permits the calculation of any averaged characteristics of an electromagnetic wave. Let $\hat{A}^{\nu}(\mathbf{r})$ be the vector potential operator ($\hbar=c=1$):

$$\hat{A}^{\nu}(\mathbf{r})=\sum_{\mathbf{k}\lambda}\frac{e^{\lambda}_{\mathbf{k}\nu}}{\sqrt{2kV}}(\hat{\alpha}_{\mathbf{k}\lambda}e^{i\mathbf{k}\cdot\mathbf{r}}+\hat{\alpha}^{+}_{\mathbf{k}\lambda}e^{-i\mathbf{k}\cdot\mathbf{r}}),$$

where $e^{\lambda}_{\mathbf{k}\nu}$ are the polarization unit vectors, $V=L_xL_yL_z$ is the normalization volume, which is assumed to be infinite in the final result, $V\to\infty$, and $\mathbf{r}$ is the observation coordinate. The mean value of the operator $\hat{A}^{\nu}(\mathbf{r})$ is found as

$$\langle\hat{A}^{\nu}(\mathbf{r})\rangle=\mathrm{Tr}\,\hat{A}^{\nu}(\mathbf{r})\rho.$$

The summation is carried out over the arguments of the electromagnetic field. In this example we demonstrate the convenience of the pseudocoordinate representation,[14] to which we are taken by the operator

$$\hat{U}=\exp(i\hat{\mathbf{k}}\mathbf{r}),\quad \hat{\mathbf{k}}=\sum_{\mathbf{k}\lambda}\mathbf{k}\hat{\alpha}^{+}_{\mathbf{k}\lambda}\hat{\alpha}_{\mathbf{k}\lambda}.$$

In the pseudocoordinate representation the operator $\hat{A}^{\nu}$ loses its dependence on $\mathbf{r}$, and the matrix

$$\rho(\mathbf{r})=\hat{U}\rho\hat{U}^{+}$$

acquires such a dependence. Now we call $\rho(\mathbf{r})$ the electromagnetic field density matrix at the point $\mathbf{r}$. Its introduction is convenient for solving boundary-value problems. In particular, if the scattering medium occupies the half-space $z>0$, to investigate the reflected field it is sufficient to know $\rho(\mathbf{r})$ at $z\to-\infty$. If there is a single photon $\hat{a}^{+}_{\mathbf{k}_0\lambda_0}|0\rangle$ in the incident mode, then[14] for $n_\mu\lambda\!\!\!^{-3}\gamma_r/\gamma<1$ the density matrix $\rho^{(c)}$ describing the reflected photon $(\mathbf{k},\lambda)$ has the form

$$\underset{z\to-\infty}{\rho^{(c)}(\mathbf{r}|1\to1)}=\sum_{\mathbf{k}}\frac{1}{8k_{0z}^4}|c^{\mathbf{k}\lambda\mathbf{k}_0\lambda_0}(k_0)|^2\times\delta(\mathbf{k},\boldsymbol{\kappa})\hat{\alpha}^{+}_{\mathbf{k}\lambda}|0\rangle\langle0|\hat{\alpha}_{\mathbf{k}\lambda},\qquad(2)$$

where

$$c^{\mathbf{k}\lambda\mathbf{k}_0\lambda_0}(k_0)=\sum_{m\mu}P^{*\lambda}_{m\mu}(\mathbf{k})P^{\lambda_0}_{m\mu}(\mathbf{k}_0)\frac{n_\mu}{k_0-\omega_{m\mu}+i\gamma/2},$$

$$\gamma=\gamma_\mu+\gamma_m,\quad \lambda\!\!\!^{-}=2\pi/k,\quad \boldsymbol{\kappa}=\{k_{0x},k_{0y},-k_{0z}\}.$$

Here we have used the two-level approximation for the atoms. In addition m labels the Zeeman sublevels of their excited state, μ labels the sublevels of the unexcited state, $\omega_{m\mu}$ is the frequency of the resonant transition, and n_μ is the concentration of scattering atoms in the μth Zeeman sublevel. The Doppler effect is assumed to be negligibly small, and γ_μ and γ_m are the collisional widths of the unexcited and excited states. It is assumed that their sum exceeds the radiation width γ_r of the excited state of the atom. Next, in the dipole approximation

$$P^{\lambda}_{m\mu}(\mathbf{k})=\frac{e}{m}\int\psi^{*}_{m}(\boldsymbol{\rho})\hat{\mathbf{p}}\mathbf{e}^{\lambda}_{\mathbf{k}}\psi_{\mu}(\boldsymbol{\rho})d\boldsymbol{\rho},\quad \hat{p}=-i\nabla,$$

where $\psi_{m\mu}$ denotes the wave functions of the valence electrons in the atoms. It is also assumed that each gas atom has one valence electron. The matrix $\rho^{(c)}(\mathbf{r}|1\to1)$ is normalized so that at $z\to-\infty$ the trace $\mathrm{Tr}\,\rho^{(c)}$ gives the probability of finding a photon in the reflection mode $(\boldsymbol{\kappa},\lambda)$. The following summation formula should be used to concretize (2):

$$\sum_{m\mu}P^{*\lambda_1}_{m\mu}(\mathbf{k}_1)P^{\lambda_2}_{m\mu}(\mathbf{k}_2)=\frac{\pi(2jm+1)}{\omega_{m\mu}}\gamma_r(\mathbf{e}^{\lambda_1}_{\mathbf{k}_1}\mathbf{e}^{\lambda_2}_{\mathbf{k}_2}).\qquad(3)$$

Here j_μ is the orbital quantum number. Then

$$\underset{z\to-\infty}{\mathrm{Tr}\,\rho^{(c)}(\mathbf{r}|1\to1)}=\frac{1}{8k_{0z}^4}\times\left|\frac{\pi(2jm+1)}{\omega_{m\mu}}\gamma_r(\mathbf{e}^{\lambda_0}_{\mathbf{k}_0}\mathbf{e}^{\lambda}_{\mathbf{k}})\frac{n_\mu(2j\mu+1)}{k_0-\omega_{m\mu}+i\gamma/2}\right|^2.\qquad(4)$$

This formula is equivalent to the Fresnel formula written for rarefied media. For the resonant frequency $k_0=\omega_{m\mu}$ we find

$$\underset{z\to-\infty}{\mathrm{Tr}\,\rho^{c}(\mathbf{r}|1\to1)}\propto\left(n_\mu\lambda\!\!\!^{-3}\frac{\gamma_r}{\gamma}\right)^2.$$

Now let the incident electromagnetic field consist of two photons which are in the same state $(\mathbf{k}_0,\lambda_0)$ and are described by a common wave function. What can we say about the distribution of the photons in the reflection mode $(\boldsymbol{\kappa},\lambda)$? We are interested in reflection at the specular angle (selective scattering). We omit the diffuse scattering. We turn to the representation (1) and examine the coherent scattering channel. We assume that under the conditions of a linear interaction between the field and individual atoms, the simultaneous scattering of two photons is mutually independent. A thin near-boundary layer of the scattering medium of thickness $\sim k_0^{-1}$ takes part in the formation of the scattered field in the coherent channel. Therefore, perturbation theory can be used. Taking into account the independent character of the scattering of the photons, for the probability of the scattering of two photons from the initial state into a two-photon final state we have

$$\underset{z\to-\infty}{\mathrm{Tr}\,\rho^{(c)}(\mathbf{r}|2\to2)}=\left|\underset{z\to-\infty}{\mathrm{Tr}\,\rho^{(c)}(\mathbf{r}|1\to1)}\right|^2\propto\left(n_\mu\lambda\!\!\!^{-3}\frac{\gamma_r}{\gamma}\right)^4.\qquad(5)$$

The calculation of $\rho^{(c)}$ is, of course, associated with the appearance of the term $\rho^{(c)}(\mathbf{r}|2\to1)$, which describes the selective scattering of only one of the two photons, while the other photon continues to propagate in the original direction of $\mathbf{k}_0$. In the lowest order of perturbation theory this term is proportional to $(n_\mu\lambda\!\!\!^{-3}\gamma_r/\gamma)^2$, and at low values of n_μ it should be decisive and have a modulus exceeding (5). This does not occur in reality. In fact, a photon propagating in the direction of $\mathbf{k}_0$ must interact with the semi-infinite

medium sooner or latter. It either replenishes the channel $\rho^{(c)}(\mathbf{r}|2\to 2)$ in the higher approximations or, if absorbed, destroys the original state of the medium and leads to the replacement of the coherent scattering channel by an incoherent channel. Thus, it is clear from physical arguments that this term should not be taken into account. Mathematically, this means that the coherent scattering channel has an infinite subsequence of terms described by Feynman diagrams in high approximations, whose sum with the term under discussion gives zero. We omit the detailed mathematical proof of this physically obvious analog of the absorption theorem.[15] Thus, the contribution to the coherent scattering channel is described only by formula (5). Two conclusions follow from the analysis performed. First, the study of the statistical properties of the scattered radiation cannot be confined to the lower orders of perturbation theory even in situations whose analysis in terms of ''quantum means'' allows perturbation theory. Second, it follows from the analog of the absorption theorem that the fate of one of the photons has an effect on the fate of the other one. Thus, the scattering of the photons cannot be represented in the form of mutually independent processes.

The validity of both conclusions is reconfirmed by an analysis of the incoherent scattering channel. In this channel a process involving the absorption of one of the photons by the medium and the elastic scattering of the other photon into the reflection mode $(\boldsymbol{\kappa},\lambda)$ should be considered in the lowest order of perturbation theory. The probability of such a process is proportional to

$$e^2 n_\mu \left(n_\mu \lambda\!\!\!^{-3} \frac{\gamma_r}{\gamma} \right)^2 |\delta(\kappa - k_0)\,\delta(k_0 - \omega_{m\mu})|^2. \tag{6}$$

This obvious result has some important consequences. It follows from it and from the relation (5) that the Poynting vector $\mathbf{s}$ of the reflected beam should be proportional to e^6 in the lowest order of perturbation theory. However, this cannot be so, because the semiclassical theory of radiation requires compliance to an $s\propto e^4$ law, according to the Heisenberg–Kramers equation.[16] The way out of this predicament is indicated by the features of the incoherent scattering channel. When the reflection coefficient is small, i.e., when $n_\mu \lambda\!\!\!^{-3}\gamma_r/\gamma<1$, the probability of the reflection of both photons must be smaller than the probability of the reflection of one of them. This means that the main role in shaping the Poynting vector is played by the incoherent channel. In this channel the power of the interaction constant e can be lowered in comparison to the first perturbative term only if the expansion coefficients are singular and the singularity grows in the high approximations. This is also the case in reality. We again arrive at the conclusion that the propagation of the photons is mutually dependent, and an infinite subsequence of Feynman diagrams must be summed.

Let us turn to the expression (6). A reasonable meaning cannot be ascribed to such a product of four Dirac δ functions. The standard way to overcome this difficulty is to interpret the combined process under investigation as two independent processes: absorption of one of the photons and scattering of the other one. The probability of each of the processes is proportional to δ^2. Such expressions can already be interpreted as the probabilities of the processes per unit time. A similar interpretation is possible, if expressions of the type (6) appear in the final formula. However, if they appear in an intermediate stage, an infinite number of terms must be taken into account in the higher approximations of perturbation theory, which ''smear'' the δ functions. For this reason, it is better to interpret the expression (6) as the scattering probability of one of the photons under the condition of absorption of the other photon and to describe this conditional probability by a single formula. The δ functions ''smeared'' by the summation of infinite subsequences no longer allow interpretation of the expression sought as a product of the probabilities of independent events. The behavior of the photons is mutually dependent. A technical procedure, which meets the task, is described below. According to the analysis performed, our confinement to the lowest approximation with respect to $n_\mu \lambda\!\!\!^{-3}\gamma_r/\gamma$ in the final formulas allows confinement to investigations of the incoherent scattering channel alone.

3. Γ-OPERATOR APPROACH

It is convenient to use the Γ-operator approach[11,12] to calculate the density matrix ρ of a photon subsystem resonantly interacting with an atomic medium with correct allowance for the photon–photon correlators. Let us consider a quantum gas, and let $\check{\psi}(\mathbf{r},t)$ be its field operator in the Heisenberg representation. Being interested in the energy density of a gas or its flux, we construct a Green's function which is bilinear with respect to the operators and completely specifies these characteristics:

$$G(x,x') = -i\langle \hat{T}\check{\psi}(x)\check{\psi}^+(x')\rangle, \quad x=\{\mathbf{r},t\}. \tag{7}$$

If we are interested in fluctuations of the energy flux, the expression (7) is inadequate. On the other hand, the function (7) specifies the particle concentration defined by the formula

$$n(x) = iG(x,x'), \quad \mathbf{r}\to\mathbf{r}', \quad t'\to t+0.$$

Now, if we study the diffusion, we can easily find any moment of the distribution:

$$\langle \mathbf{r}^m\rangle = \frac{\int \mathbf{r}^m G(x,x')\,d\mathbf{r}}{\int G(x,x')\,d\mathbf{r}}, \quad \mathbf{r}\to\mathbf{r}', \quad t'\to t+0.$$

Thus, the Green's function plays the role of the distribution function of parameters which are arguments of the averaged field operators $\check{\psi}$ and $\check{\psi}^+$. Under second quantization the arguments (the coordinates of the particles) of the wave function $\psi(\mathbf{r},t)$ become arguments of the wave operators $\check{\psi}(\mathbf{r},t)$. The use of quantum Green's functions in the second quantization formalism makes it possible to calculate the distribution function (density matrix) of the particles as a function of their coordinates. The transformation scheme is as follows:

$$\psi(\mathbf{r},t)\to\check{\psi}(\mathbf{r},t)\to iG(x,x')\to n(x). \tag{8}$$

The quantized electromagnetic field is described by the wave function $\Phi(\mathbf{N},t)$, where the vector $\mathbf{N}=\ldots,N_{\mathbf{k}\lambda},\ldots$ is the set of the occupation numbers $N_{\mathbf{k}\lambda}$ of the individual modes

$(\mathbf{k},\lambda)$. We are interested in the density matrix ρ, which depends on $\mathbf{N}$. In analogy to (8), the calculation scheme takes the form

$$\Phi(\mathbf{N},t)\to\check{\Phi}(\mathbf{N},t)\to i\mathscr{D}(\mathbf{N},t,\mathbf{N}',t')\to\rho.$$

Thus, if we wish to know $\rho(\mathbf{N},t)$, we can formally proceed in the following manner. The second-quantized system described by the function $\Phi(\mathbf{N},t)$ must be quantized once again. The operator $\check{\Phi}(\mathbf{N},t)$ appears as a result. The Green's function $\mathscr{D}$ should be constructed in the formalism thus created. Actually, the formalism we used, which is reminiscent of the ''third'' quantization of fields, is essentially different. The ''third'' quantization that we used is a formal operation and is constructed using a specially selected unitary transformation of the dynamic equations, which does not deform the fundamental principles. We again refer to an analogy. The quantum mechanics of a single particle can be constructed in the coordinate representation on the basis of the wave function $\psi(\mathbf{r},t)$, or the second-quantization representation can be used. The relationship between them is unitary. In the second-quantization representation the Bose and Fermi fields lead to coinciding results in the case of only one particle. This remark is significant for devising our formalism. The use of a unitary transformation allows us to abandon the term ''third'' quantization and to call the representation constructed a Γ representation in analogy to the many-particle Γ space of a statistical function. Thus, from the formal standpoint, the construction of the Γ representation from the second-quantization representation actually imitates the method for constructing the second-quantization representation from the configuration representation.[18] For this reason, any discord in the Γ representation would at once indicate similar discord in the second-quantization representation. On the other hand, all the investigative methods developed for quantum systems, such as Feynman diagrams, Dyson equations, spectral representations, dispersion relations, etc., can be used in the Γ representation.

Let the second-quantized transverse electromagnetic field interact quasiresonantly with a gas consisting of atoms with one valence electron. We write the Schrödinger equation of the system in the form

$$i\frac{\partial\Psi}{\partial t}=\hat{H}\Psi,\quad \hat{H}=\hat{H}_0+\hat{H}',\quad \hat{H}_0=\hat{H}_{\rm ph}+\hat{H}_a,$$

$$\hat{H}_{\rm ph}=\sum_{\mathbf{k}\lambda} k\hat{\alpha}^+_{\mathbf{k}\lambda}\hat{\alpha}_{\mathbf{k}\lambda},\quad \hat{H}_a=\sum_{i\mathbf{p}}\varepsilon_i(\mathbf{p})\hat{b}^+_{i\mathbf{p}}\hat{b}_{i\mathbf{p}},$$

$$\hat{H}'=-\frac{e}{m}\int\hat{\psi}^+\hat{\mathbf{p}}\hat{\mathbf{A}}(\mathbf{r})\hat{\psi}d\mathbf{r}d\mathbf{R}+\tilde{H},$$

$$\hat{\psi}(\mathbf{r},\mathbf{R})=\sum_{i\mathbf{p}}\psi_i(\mathbf{r}-\mathbf{R})\frac{e^{i\mathbf{p}\cdot\mathbf{R}}}{\sqrt{V}}\hat{b}_{i\mathbf{p}},\quad \varepsilon_i(\mathbf{p})=\varepsilon_i+\frac{p^2}{2M}.$$

Here $\psi_i(\mathbf{r}-\mathbf{R})$ are the wave functions of electrons in atoms with an energy ε_i, $\mathbf{r}$ and $\mathbf{R}$ are the coordinates of the electron and the remainder of the atom, $\mathbf{p}$ is the momentum of the atom, and M is its mass. Also, $\tilde{H}$ denotes the Hamiltonian of the interaction of gas atoms with the reservoir. In the absence of temperature degeneracy of the atoms, which is implied, the annihilation (creation) operators $\hat{b}_{i\mathbf{p}}$ ($\hat{b}^+_{i\mathbf{p}}$) of atoms in the state $(i,\mathbf{p})$ can be assumed to belong to the Bose–Einstein field. We omit taking into account the longitudinal component of the electromagnetic field, understanding that in media of elevated density it reveals itself by complicating the effects predicted here without eliminating them.

To calculate $\rho(\mathbf{N},t)$ we use the Γ-operator approach, which automatically separates the coherent and incoherent scattering channels. We construct the Γ representation in the following manner. Any state of a free second-quantized electromagnetic field with a definite energy can be described by the wave function

$$\Phi^0(\mathbf{N}|\boldsymbol{\zeta})=\prod_{\mathbf{k}\lambda}\varphi(N_{\mathbf{k}\lambda}|\zeta_{\mathbf{k}\lambda}),\tag{9}$$

where φ is the wave function of the quantum oscillator, which depends on the arguments $\zeta_{\mathbf{k}\lambda}$. The physical meaning of these arguments is not important to us at this point. We construct a Γ space with the creation vector $\rangle^0_\Gamma$ (Refs. 11 and 12) and stipulate that this vector is a mathematical vacuum. Let $\hat{\mathfrak{A}}(\mathbf{N})$ be the annihilation operator of an ensemble of noninteracting photons with the set of occupation numbers $\mathbf{N}$, and let $\hat{\mathfrak{A}}^+(\mathbf{N})\rangle^0_\Gamma$ be the wave function of such a state. These wave functions form a complete basis for expanding any wave function describing physically permissible states. Between the basis vectors $\hat{\mathfrak{A}}^+(\mathbf{N})\rangle^0_\Gamma$ thus introduced and the functions (9) there is a unitary transformation, which is effected by the operator $\hat{O}$:[11,12]

$$\hat{O}=\hat{\Phi}^+\rangle^0_\Gamma,\quad \hat{\Phi}(\boldsymbol{\zeta})=\sum_{\mathbf{N}}\Phi^0(\mathbf{N}|\boldsymbol{\zeta})\hat{\mathfrak{A}}(\mathbf{N}),\quad \Psi_\Gamma=\hat{O}\Psi.$$

If it is assumed that the operators $\hat{\mathfrak{A}}(\mathbf{N})$ and $\hat{\mathfrak{A}}^+(\mathbf{N})$ obey the commutation relations for a Bose–Einstein field

$$[\hat{\mathfrak{A}}(\mathbf{N}),\hat{\mathfrak{A}}^+(\mathbf{N}')]=\delta(\mathbf{N},\mathbf{N}'),\tag{10}$$

it is clear that the Γ space is isomorphous to the space of occupation numbers of the Bose particles. However, the occupation numbers in the Γ space do not exceed unity, since no physical state corresponds to the vector $(\hat{\mathfrak{A}}^+(\mathbf{N}))^n\rangle^0_\Gamma$ when $n>1$. It can be stated that there is one ''superparticle,'' for which the basis $\hat{\mathfrak{A}}^+(\mathbf{N})\rangle^0_\Gamma$ is sufficient for describing its kinetics, in the Γ space. However, the behavior of one particle can be described with equal success using Bose and Fermi field operators. This means that anticommutation relations can be used for $\hat{\mathfrak{A}}$ and $\hat{\mathfrak{A}}^+$. The result of the calculations remains unchanged. Other commutation relations are also permissible for them.[19] Hence it is seen that these operators have no physical meaning. For simplicity, we shall use the commutation relation (10) for them. In the Γ representation the Schrödinger equation takes the form[12]

$$i\frac{\partial\Psi_\Gamma}{\partial t}=\left[\hat{H}_a+\sum_{\mathbf{N}}\varepsilon(\mathbf{N})\hat{\mathfrak{A}}^+(\mathbf{N})\hat{\mathfrak{A}}(\mathbf{N})-\frac{e}{m}\int\hat{\Phi}^+\hat{\psi}^+\hat{\mathbf{p}}\hat{\mathbf{A}}(\mathbf{r})\hat{\psi}\hat{\Phi}d\mathbf{r}d\mathbf{R}d\boldsymbol{\zeta}\right]\Psi_\Gamma+\tilde{H}\Psi_\Gamma,$$

where

$$\varepsilon(\mathbf{N})=\sum_{\mathbf{k}\lambda} kN_{\mathbf{k}\lambda}, \quad d\boldsymbol{\zeta}=\prod_{\mathbf{k}\lambda} d\zeta_{\mathbf{k}\lambda}.$$

The mean value of any photon operator $\hat{K}$ can be calculated using the formula[12]

$$\langle\hat{K}\rangle=\mathrm{Tr}\,\hat{K}\rho=\int \hat{K}(\boldsymbol{\zeta}',\boldsymbol{\zeta})\rho(\boldsymbol{\zeta},\boldsymbol{\zeta}')d\boldsymbol{\zeta}d\boldsymbol{\zeta}', \tag{11}$$

where

$$\rho(\boldsymbol{\zeta},\boldsymbol{\zeta}')=\langle\hat{\Phi}^{+}(\boldsymbol{\zeta}')\hat{\Phi}(\boldsymbol{\zeta})\rangle_{\Gamma}, \quad \rangle_{\Gamma}=\Psi_{\Gamma}. \tag{12}$$

The expression (11) is valid for any operator $\hat{K}$. This means that the expression (12) is the density matrix sought ρ for the photonic subsystem interacting with an atomic gas. To calculate ρ we utilize the kinetic Green's function approach[20] in the Γ representation. We introduce

$$\mathscr{D}_{ll'}(\boldsymbol{\zeta},t,\boldsymbol{\zeta}',t')=-i\langle\hat{T}_c\tilde{\Phi}_l(\boldsymbol{\zeta},t)\tilde{\Phi}^{+}_{l'}(\boldsymbol{\zeta}',t')\hat{S}_c\rangle_{0\Gamma}, \tag{13}$$

where $\tilde{\Phi}$ is a field operator in the interaction representation:

$$\tilde{\Phi}(\zeta,t)=\sum_{\mathbf{N}} \Phi^{0}(\mathbf{N}|\boldsymbol{\zeta})\hat{\mathfrak{A}}(\mathbf{N})\exp[-i\varepsilon(\mathbf{N})t].$$

The averaging in (13) is carried out over the initial state of the atomic system and the field of the photons before the interaction between them occurs. The label l describes the temporal contour, which originates ($l=1$) at $t\to-\infty$, extends to $t\to\infty$, and returns back ($l=2$) to $t\to-\infty$, and $\hat{T}_c$ is the chronological operator on that contour. The operator $\hat{S}_c$ has the form

$$\hat{S}_c=\hat{T}_c\exp\Bigg\{\sum_{l}(-1)^{l+1}\frac{ie}{m}\times\int\tilde{\Phi}^{+}_{l}\tilde{\psi}^{+}_{l}\hat{\mathbf{p}}\hat{\mathbf{A}}(\mathbf{r})\tilde{\psi}_{l}\tilde{\Phi}_{l}d\boldsymbol{\zeta}\,d\mathbf{r}\,d\mathbf{R}\,dt\Bigg\},$$

$$\tilde{\psi}=\sum_{i\mathbf{p}}\psi_{i}(\mathbf{r}-\mathbf{R})\frac{\hat{b}_{i\mathbf{p}}}{\sqrt{V}}\exp[i\mathbf{p}\cdot\mathbf{R}-i\varepsilon_{i}(p)t].$$

We have omitted the operator $\tilde{H}$, whose influence is manifested through the mass operators by broadening of the energy levels of the atoms. The standard technique for going over to a normal product of field operators is used to calculate (13). However, since[11] $(\hat{\mathfrak{A}}(\mathbf{N}))^{n}\rangle_{\Gamma}=0$ in physical states when $n>1$, all the normal products, except the simplest, give a zero result. This is the source of the main advantage of the Γ-operator formalism. The functions $\mathscr{D}_{ll'}$ have the following structure:[11]

$$\mathscr{D}_{ll'}=\Delta_{ll'}-i\rho_{ll'}.$$

Thus, $\rho=\rho_{12}=i\mathscr{D}_{12}$ at $t=t'$. For $\Delta_{ll'}$ and $\rho_{ll'}$ we have the following system of equations:

$$\rho_{12}=\rho^{(c)}_{12}+\rho^{(n)}_{12}, \quad \rho^{(c)}_{12}=(1+\Delta_r\hat{\mathscr{P}}_r)\rho^{0}_{12}(1+\hat{\mathscr{P}}_a\Delta_a), \tag{14}$$

$$\rho^{(n)}_{12}=-\Delta_r\hat{\mathscr{P}}^{(n)}_{12}\Delta_a, \quad \Delta_r=\Delta^{0}_{r}+\Delta^{0}_{r}\hat{\mathscr{P}}_r\Delta_r.$$

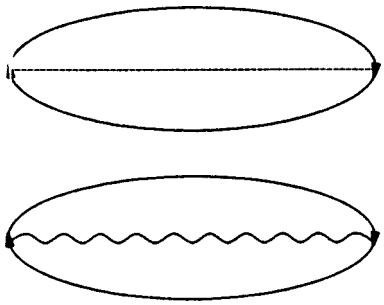

FIG. 1. Schematic representation of the polarization operators $\hat{\mathscr{P}}_{l_1l_2}$ and $\hat{\mathscr{P}}^{(n)}_{l_1l_2}$ corresponding to formulas (15). Solid lines—the atomic functions $G_{l_1l_2}$, dashed line—the propagator $\Delta^{0}_{l_1l_2}$, wavy line—the matrix $\rho^{0}_{l_1l_2}$.

Here $\hat{\mathscr{P}}_{ll'}$ and $\hat{\mathscr{P}}^{(n)}_{ll'}$ are the polarization operators in the Γ formalism,

$$\Delta_r=\Delta_{11}, \quad \Delta_a=-\Delta_{22}=\Delta^{+}_{r},$$

$$\hat{\mathscr{P}}_r=\hat{\mathscr{P}}_{11}, \quad \hat{\mathscr{P}}_a=\hat{\mathscr{P}}_{22}=\hat{\mathscr{P}}^{+}_{r},$$

and ρ^{0} is the density matrix of the free field of the photons. In the energy representation

$$\Delta^{0}_{r}=(E-\hat{H}_{\mathrm{ph}}+i0)^{-1}.$$

The simple polarization operators

$$\hat{\mathscr{P}}_{l_1l_2}=-\left(\frac{e}{m}\right)^{2}\sum_{\nu_1\nu_2}\int\hat{p}^{\nu_1}\hat{A}^{\nu_1}G^{0}_{l_1l_2}(X_1,X_2)\Delta^{0}_{l_1l_2}(-1)^{l_2+1}\times\hat{p}^{\nu_2}\hat{A}^{\nu_2}G^{0}_{l_2l_1}(X_2,X_1)d\mathbf{r}_1d\mathbf{r}_2d\mathbf{R}_1d\mathbf{R}_2,$$

$$X=\{\mathbf{r},\mathbf{R},t\},$$

$$\hat{\mathscr{P}}^{(n)}_{l_1l_2}=-\left(\frac{e}{m}\right)^{2}\sum_{\nu_1\nu_2}\int\hat{p}^{\nu_1}\hat{A}^{\nu_1}G^{0}_{l_1l_2}(X_1,X_2)\times\rho^{0}_{l_1l_2}(-1)^{l_2+1}\hat{p}^{\nu_2}\hat{A}^{\nu_2}G^{0}_{l_2l_1}(X_2,X_1)d\mathbf{r}_1d\mathbf{r}_2d\mathbf{R}_1d\mathbf{R}_2 \tag{15}$$

are represented by the Feynman graphs shown in Fig. 1. In this figure the solid lines depict the functions $G^{0}_{l_1l_2}$, the dashed line represents $\Delta^{0}_{l_1l_2}$, and the wavy line corresponds to $\rho^{0}_{l_1l_2}$. Each vertex is associated with a multiplier $\hat{p}^{\nu}\hat{A}^{\nu}$. We borrow the explicit form of the Green's function of the atomic system from Ref. 14.

4. INCOHERENT SCATTERING CHANNEL

We are interested in the probability of finding one photon in the reflection mode $(\boldsymbol{\kappa},\lambda)$, if the occupation number of photons in the mode $(\mathbf{k}_0,\lambda_0)$ is equal to two. In the lowest order of perturbation theory the corresponding matrix $\rho^{(n)}(\mathbf{r}|2\to1)$ is associated with the sum of Feynman diagrams depicted in Fig. 2. We demonstrate the calculation of these diagrams in the case of the first of them. The analytical expression corresponding to it follows from iterations of the system of equations (14):

$$\rho^{(n)}_{12}(\mathbf{r}|2\to1)=-\Delta^{0}_{r}\hat{\mathscr{P}}_r\Delta^{0}_{r}\hat{\mathscr{P}}^{(n)}_{12}\Delta^{0}_{a}\hat{\mathscr{P}}_a\Delta^{0}_{a}. \tag{16}$$

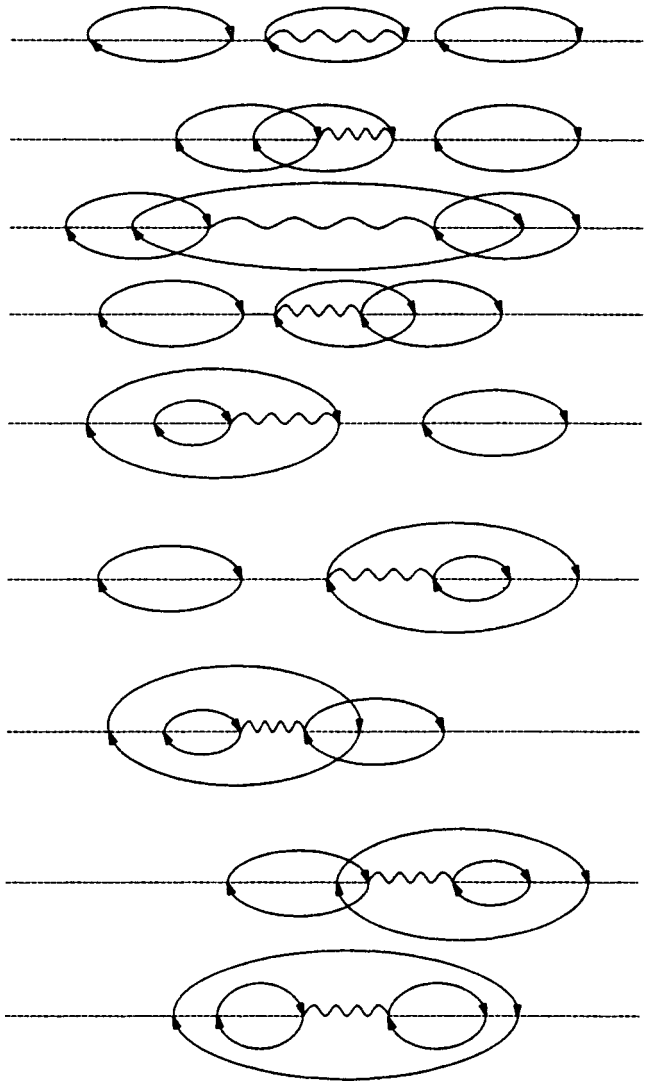

FIG. 2. Diagrams of incoherent scattering processes supplying one photon to the reflection mode when there are two photons in the incident mode.

According to the rules adopted, the left-hand function Δ_r^0 in (16) is represented in the first diagram in Fig. 2 by the left-hand dashed line. The next structure in this diagram is a representation of the operator $\hat{\mathscr{P}}_r$, which is consistent with Fig. 1. It is followed by a dashed line corresponding to Δ_r^0 and then by the representation of the operator $\hat{\mathscr{P}}_{12}^{(n)}$. The right-hand part of the graph is constructed in a symmetrical manner in accordance with the analytical expression (16). Formulas (15) are used for $\hat{\mathscr{P}}_r$ and $\hat{\mathscr{P}}_{12}^{(n)}$. If the space is homogeneous and the scattering process is stationary, in the energy representation[11] we have

$$\hat{\mathscr{P}}_r^{\infty}(E)=\sum_{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}\hat{\alpha}^{+}_{\mathbf{k}_1\lambda_1}C_r^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E-\hat{H}_{\rm ph})\hat{\alpha}_{\mathbf{k}_2\lambda_2},$$

$$C_r^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E)=c_r^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E)\,\delta(\mathbf{k}_1,\mathbf{k}_2), \tag{17}$$

$$c_r^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E)=\sum_{m\mu}\frac{P^{*\lambda_1}_{m\mu}(\mathbf{k}_1)P^{\lambda_2}_{m\mu}(\mathbf{k}_2)}{2kV}\,\frac{N_\mu}{E-\omega_{m\mu}+i\gamma/2}.$$

As before, we neglect the Doppler effect, and N_μ is the occupation number of the μth sublevel of the atoms. Using the summation rule (3), for $\mathrm{Im}\,\hat{\mathscr{P}}_r^{\infty}$ we obtain

$$\mathrm{Im}\,\hat{\mathscr{P}}_r^{\infty}(E)\hat{\alpha}_{\mathbf{k}\lambda}|0\rangle=-\frac{1}{2}\gamma_{\rm ph}(E)\hat{\alpha}^{+}_{\mathbf{k}\lambda}|0\rangle,$$

$$\gamma_{\rm ph}(E)=\frac{\pi^2(2j_m+1)}{\omega_{m\mu}^2}\gamma_r n_\mu(2j_\mu+1)\,\delta_\gamma(E-\omega_{m\mu}), \tag{18}$$

$$\delta_\gamma(E)=-\left[\left(E+\frac{i\gamma}{2}\right)^{-1}-\left(E-\frac{i\gamma}{2}\right)^{-1}\right]\frac{1}{2\pi i}.$$

For $\hat{\mathscr{P}}_{12}^{(n)}$ from (15) we obtain

$$\hat{\mathscr{P}}_{12}^{(n)}(E)=\sum_{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}\hat{\alpha}_{\mathbf{k}_1\lambda_1}\int C_{12}^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E_1)\times\rho_{12}(E-E_1)\frac{dE_1}{2\pi}\hat{\alpha}^{+}_{\mathbf{k}_2\lambda_2}.$$

If the medium is spatially homogeneous, then

$$C_{12}^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E)=\delta(\mathbf{k}_1,\mathbf{k}_2)c_{12}^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E),$$

$$c_{12}^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}(E)=-2\pi\sum_{m\mu}\frac{P^{\lambda_1}_{m\mu}(\mathbf{k}_1)P^{*\lambda_2}_{m\mu}(\mathbf{k}_2)}{2kV}\times N_\mu\delta_\gamma(E+\varepsilon_m-\varepsilon_\mu). \tag{19}$$

If the scattering medium occupies the half-space $z>0$, then in the Wigner approximation the following replacement must be made in formulas (16) and (18):[11]

$$\delta(\mathbf{k}_1,\mathbf{k}_2)\to\delta(k_{1x},k_{2x})\,\delta(k_{1y},k_{2y})\,\delta^{+}(k_{1z}-k_{2z}),$$

where

$$\delta^{+}(q)=\int_0^{L_z/2}e^{-iqz}\frac{dz}{L_z}.$$

Taking into account the difference dependence of $\rho_{12}^{(n)}$ on t and t' under stationary conditions, after Fourier transformation, in agreement with formula (16), we have

$$\begin{aligned}\rho_{12}^{(n)}(E|2\to1)=&-(E-\hat{H}_{\rm ph}+i0)^{-1}\\&\times\sum_{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}\hat{\alpha}^{+}_{\mathbf{k}_1\lambda_1}C_r^{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}\\&\times(E-\hat{H}_{\rm ph})\hat{\alpha}_{\mathbf{k}_2\lambda_2}(E-\hat{H}_{\rm ph}+i0)^{-1}\\&\times\sum_{\mathbf{k}_3\lambda_3\mathbf{k}_4\lambda_4}\hat{\alpha}_{\mathbf{k}_3\lambda_3}C_{12}^{\mathbf{k}_3\lambda_3\mathbf{k}_4\lambda_4}\rho_{12}^{0}\hat{\alpha}^{+}_{\mathbf{k}_4\lambda_4}\\&\times(E-\hat{H}_{\rm ph}-i0)^{-1}\sum_{\mathbf{k}_5\lambda_5\mathbf{k}_6\lambda_6}\hat{\alpha}^{+}_{\mathbf{k}_5\lambda_5}C_a^{\mathbf{k}_5\lambda_5\mathbf{k}_6\lambda_6}\\&\times(E-\hat{H}_{\rm ph})\hat{\alpha}_{\mathbf{k}_6\lambda_6}(E-\hat{H}_{\rm ph}-i0)^{-1}.\end{aligned} \tag{20}$$

In the pseudocoordinate representation the expression (20) takes the form

$$\rho_{12}^{(n)}(\mathbf{r},E|2\to 1)=-\sum_{\mathbf{k}_1\lambda_1\mathbf{k}_2\lambda_2}\frac{\exp(i\mathbf{k}_1\cdot\mathbf{r})}{E-k_1+i0}\frac{C_r^{\mathbf{k}_1\lambda_1\mathbf{k}_0\lambda_0}(E)}{E-k_0+i0}$$
$$\times\frac{C_{12}^{\mathbf{k}_0\lambda_0\mathbf{k}_0\lambda_0}(E-2k_0)}{E-k_0-i0}C_a^{\mathbf{k}_0\lambda_0\mathbf{k}_2\lambda_2}(E)$$
$$\times\frac{\exp(-i\mathbf{k}_2\cdot\mathbf{r})}{E-k_2-i0}\hat{\alpha}^+_{\mathbf{k}_1\lambda_1}|0\rangle\langle 0|\hat{\alpha}_{\mathbf{k}_2\lambda_2}.$$

In this case

$$\rho_{12}^{0}(E)=\pi(\hat{\alpha}^+_{\mathbf{k}_0\lambda_0})^2|0\rangle\langle 0|(\hat{\alpha}_{\mathbf{k}_0\lambda_0})^2\delta(E-2k_0).$$

Since we are interested only in the reflected beam, we set $z\to-\infty$. These circumstances permit the use of the asymptotic formula[14]

$$\frac{1}{2k}\frac{\exp(ik_zz)}{E-k+i0}\xrightarrow[z\to-\infty]{}iL_z\frac{\exp(ik_zz)}{2k_z}$$
$$\times\delta(k_z,-\sqrt{E^2-k_x^2-k_y^2})\xrightarrow[V\to\infty]{}2\pi i\frac{\exp(ik_zz)}{2k_z}$$
$$\times\delta(k_z+\sqrt{E^2-k_x^2-k_y^2}). \tag{21}$$

Here $\delta(k,k')$ is a Kronecker delta, and $\delta(k-k')$ is a Dirac delta function. The structure of formula (21) ensures correct results following the subsequent limiting transition $V\to\infty$ and performance of the operation $\mathrm{Tr}\,\hat{K}\rho_{12}$ with any operator $\hat{K}$. The product of Kronecker deltas

$$\delta(k_{1z},-\sqrt{E^2-k_{1x}^2-k_{1y}^2})\,\delta(k_{2z},-\sqrt{E^2-k_{2x}^2-k_{2y}^2})$$

can be rewritten by virtue of the equalities $k_{1x}=k_{2x}$ and $k_{1y}=k_{2y}$ in the form

$$\delta(k_{1z},k_{2z})\,\delta(k_{1z},-\sqrt{E^2-k_{1x}^2-k_{1y}^2}).$$

Bearing in mind the subsequent limiting transition $V\to\infty$, we write, in agreement with (21),

$$\delta(k_{1z},-\sqrt{E^2-k_{1x}^2-k_{1y}^2})\to\frac{2\pi}{L_z}$$
$$\times\delta(k_{1z}+\sqrt{E^2-k_{1x}^2-k_{1y}^2}).$$

Now we are in a position to carry out the integration over E after performing the replacement of variables

$$E^2-k_{1x}^2-k_{1y}^2=t^2.$$

We then have

$$\rho_{12}^{(n)}(\mathbf{r}|2\to 1)=\int\rho_{12}^{(n)}(\mathbf{r},E|2\to 1)\frac{dE}{2\pi}$$
$$=L_z\sum_{\mathbf{k}_1\lambda_1k_2\lambda_2}\frac{k_1}{k_{1z}}\frac{C_r^{\mathbf{k}_1\lambda_1\mathbf{k}_0\lambda_0}(k_1)}{k_1-k_0+i0}$$
$$\times C_{12}^{\mathbf{k}_0\lambda_0\mathbf{k}_0\lambda_0}(k_1-2k_0)$$
$$\times\frac{C_a^{\mathbf{k}_0\lambda_0\mathbf{k}_2\lambda_2}(k_1)}{k_1-k_0-i0}\hat{\alpha}^+_{\mathbf{k}_1\lambda_1}|0\rangle\langle 0|\hat{\alpha}_{\mathbf{k}_1\lambda_2}. \tag{22}$$

The expression (22) has a nonintegrable pole at $k_1=k_0$. This pole is the feature which was mentioned in Sec. 2.

The appearance of this singularity is caused by the propagators $\Delta_{r,a}^0$ in (20), which describe the evolution of a photon in the medium. The singularity vanishes, if the complete propagators, which take into account the interaction of the photon with the medium, are used instead of $\Delta_{r,a}^0$. As an acceptable approximation, we replace $\Delta_{r,a}^0$ by the propagators of a photon in an infinite medium $\Delta_{r,a}^\infty$ (Ref. 12):

$$\Delta_r^0(E)\to\Delta_r^\infty(E),$$
$$\Delta_r^\infty(E)\,\hat{\alpha}^+_{\mathbf{k}\lambda}|0\rangle=\left(E-k+i\frac{\gamma_{\mathrm{ph}}}{2}\right)\hat{\alpha}^+_{\mathbf{k}\lambda}|0\rangle;$$
$$\Delta_a^0(E)\to\Delta_a^\infty(E),$$
$$\Delta_a^\infty(E)\,\hat{\alpha}^+_{\mathbf{k}\lambda}|0\rangle=\left(E-k-i\frac{\gamma_{\mathrm{ph}}}{2}\right)\hat{\alpha}^+_{\mathbf{k}\lambda}|0\rangle.$$

This replacement leads to the following replacement in (22) as a result:

$$|(k_1-k_0+i0)^{-1}|^2\to\left|\left(k_1-k_0+i\frac{\gamma_{\mathrm{ph}}(k_1)}{2}\right)^{-1}\right|^2.$$

Then

$$k_1k_0C_r^{\mathbf{k}_1\lambda_1\mathbf{k}_0\lambda_0}(k_1)C_a^{\mathbf{k}_0\lambda_0\mathbf{k}_1\lambda_1}(k_1)$$
$$=\frac{\sin^2[(k_{1z}-k_{0z})L_z/4]}{(k_{1z}-k_{0z})^2L_z^2}\delta(k_{1x},k_{0x})\,\delta(k_{1y},k_{0y})$$
$$\times\delta(k_{2x},k_{0x})\,\delta(k_{2y},k_{0y})|c^{\mathbf{k}_1\lambda_1\mathbf{k}_0\lambda_0}(k_0)|^2.$$

Since we are interested in the diagonal elements of the density matrix, we set $\lambda_1=\lambda_2$. Then formula (22) can be rewritten for $\gamma\gg\gamma_{\mathrm{ph}}$ in the form ($L_z\to\infty$)

$$\rho_{12}^{(n)}(\mathbf{r}|2\to 1)=\frac{1}{L_z}\sum_{\mathbf{k}_1\lambda_1}\frac{\pi^2(2j_m+1)(2j_\mu+1)k_{1z}}{\omega_{m\mu}^2k_1}\gamma_rn_\mu$$
$$\times\delta_\gamma(\omega_{m\mu}-k_0)\frac{\delta(k_{1x},k_{0x})\,\delta(k_{1y},k_{0y})}{(k_1-k_0)^2+\gamma_{\mathrm{ph}}^2(k_0)/4}$$
$$\times\hat{\alpha}^+_{\mathbf{k}_1\lambda_1}|0\rangle\langle 0|\hat{\alpha}_{\mathbf{k}_1\lambda_1}\mathrm{Tr}\,\rho_{12}^{(c)}(\mathbf{r}|1\to 1). \tag{23}$$

Because of the presence of γ and γ_{ph} in the denominators in (23), the equalities $k_{1z}=-k_{0z}$ and $k_1=k_0$ do not hold in the reflected flux. Thus, we have a sharp frequency-angle distribution in the reflected flux with a half-width proportional to

$$\Delta\vartheta\propto\left(\sqrt{\cos\vartheta+\frac{\gamma_{\mathrm{ph}}(k_0)}{k_0}}-\cos\vartheta\right)\sin\vartheta,$$

where ϑ is the angle of incidence of the beam. A similar frequency-angle distribution was previously discovered[12] in an investigation of the reflection of a one-photon state from a thermally excited medium. The semiclassical theory of radiation does not predict such a distribution. This distribution likewise does not appear in calculations based on the Green's

function approach,[17] which provides for a discontinuity in the photon–photon correlators. Thus, the presence of a frequency-angle distribution in the direction of specular reflection is a characteristic consequence of the mutual correlation of photons in the reflection process.

There should be interest in the integral probability of the reflection of one photon of polarization λ, i.e.,

$$\operatorname{Tr}\rho_{12}^{(n)}(\mathbf{r}|2\to 1).$$
$$z\to-\infty$$

Here we have omitted the summation over λ.

After the sums in (23) vanish because of the presence of the Kronecker deltas, the summation over k_{1z} remains. As $L_z\to\infty$, it is replaced by integration, and for $\gamma\gg\gamma_{\rm ph}$ we have

$$\frac{\operatorname{Tr}\rho_{12}^{(n)}(\mathbf{r}|2\to 1)}{\operatorname{Tr}\rho_{12}^{(c)}(\mathbf{r}|1\to 1)}=-\frac{\pi(2j_m+1)}{2\omega_{m\mu}^2}\times\gamma_r n_\mu\delta_\gamma(\omega_{m\mu}-k_0)(2j_\mu+1)\times\int_{-\infty}^{0}\frac{k_{1z}}{(k_1-k_0)^2+\gamma_{\rm ph}^2(k_0)/4}\frac{dk_{1z}}{k_1}. \quad (24)$$

In the latter integral the vicinity of $k_{1z}=-k_{0z}$ is significant. For this reason we use the expansion

$$k_1=k_0-\frac{k_{0z}}{k_0}(k_{1z}+k_{0z})$$

and replace the upper integration limit by infinity. With consideration of (18) the expression (24) is equal to unity. Now it should be recalled that we calculated the contribution of only the first diagram depicted in Fig. 2. Consideration of the other diagrams leads to the following replacement in (23):

$$\left[(k_1-k_0)^2+\frac{\gamma_{\rm ph}^2}{4}\right]^{-1}\to 4\pi^2\delta_{\gamma_{\rm ph}}^2(k_1-k_0).$$

In response to this replacement, the value of the ratio (24) doubles. As a result, we have

$$\frac{\operatorname{Tr}\rho_{12}^{(n)}(\mathbf{r}|2\to 1)}{\operatorname{Tr}\rho_{12}^{(c)}(\mathbf{r}|1\to 1)}=2,\quad z\to-\infty. \quad (25)$$

The two corresponds to the number of photons in the incident mode. Since the value of $\rho^{(c)}(\mathbf{r}|1\to 1)$ at $z\to-\infty$ specifies the probability of the reflection of a photon when there is one photon in the original mode $(\mathbf{k}_0,\lambda_0)$, the integral result can be interpreted as a consequence of the independent scattering of the photons. The correlation properties of the photons are moderated in the integral characteristic (25). Formula (25) is formally reminiscent of formula (28) from Ref. 12. In Ref. 12 the medium was assumed to be excited, and the scattering processes were, in a certain sense, the reverses of the processes studied here. The difference by a factor of 2 is caused by the fact that in Ref. 12 the expression for the operator $\hat{\mathscr{T}}_{12}^{(n)}$ in a homogeneous space was used as an approximation to calculate the parameters of the radiation reflected from the excited medium. Consideration of the finite dimensions of the medium in this operator alters the result by a factor of 2.

In agreement with formulas (5) and (25) at $z\to-\infty$, the integral reflection coefficient R for $n_\mu\lambda^3\gamma_r/\gamma<1$ equals

$$R=\operatorname{Tr}\rho_{12}^{(n)}(\mathbf{r}|2\to 1)/\operatorname{Tr}(\hat{\alpha}_{\mathbf{k}_0\lambda_0}^{+}\hat{\alpha}_{\mathbf{k}_0\lambda_0}\rho^0)$$
$$=\operatorname{Tr}\rho^{(c)}(\mathbf{r}|1\to 1),\quad z\to-\infty.$$

Thus, we return to formula (4), i.e., to the classical Fresnel formula, which is analytically dependent on the interaction constant of the field with matter as $e\to 0$. Only in calculating such an integral characteristic as R can the photons be considered mutually independent and can the concept of the refractive index of the medium, which is what specifies the Fresnel formulas, be used. We recall that before integration over the angle ϑ, we had the dependence (23), which is not analytic with respect to e, for specifying the frequency-angle distribution of the photons. The refractive index is insufficient for describing such a frequency-angle distribution.

The calculation result is different, if $\gamma\ll\gamma_{\rm ph}$. This condition can easily be realized in an experiment. Instead of Eq. (25), when $k_0=\omega_{m\mu}$, we have the following relation:

$$\frac{\operatorname{Tr}\rho_{12}^{(n)}(\mathbf{r}|2\to 1)}{\operatorname{Tr}\rho_{12}^{(c)}(\mathbf{r}|1\to 1)}\propto\frac{\gamma}{\gamma_{\rm ph}(\omega_{m\mu})},\quad z\to-\infty. \quad (26)$$

Under these conditions the Fresnel formulas are violated, and reflection is suppressed. In other words, consideration of the mutual correlation of photons for $\gamma\ll\gamma_{\rm ph}$ leads to inapplicability of the semiclassical theory of radiation for calculating the reflection of resonance radiation even from unexcited media as long as the photons are mutually correlated in the incident flux. In this case the reflection coefficient R begins to depend on the statistical properties of the incident radiation even when the interaction of light with the individual atoms of the medium is linear.

The suppression of the reflection of mutually correlated photons is subject to experimental testing. The quantum structure of the electromagnetic flux in free space can, in fact, be different. While the photons in the radiation flux from a laser are mutually correlated for the most part, the fraction of such photons in a flux of black-body radiation is considerably smaller. Thus, in accordance with formula (26) we should expect a decrease in the reflection coefficient R when a thermal source of a incident flux is replaced by a laser source, which can be detected by comparing the critical concentrations of atoms that determined the mutual switching between the diffuse and specular types of reflection in Wood's experiment[6] when thermal and laser sources of incident radiation are employed.

*)E-mail: phf@deans.mpei.ac.ru

[1] Yu. K. Zemtsov, A. Yu. Sechin, A. N. Starostin, A. G. Leonov, A. A. Rudenko, and D. I. Chekhov, Zh. Éksp. Teor. Fiz. **114**, 135 (1998) [JETP **87**, 76 (1998)].
[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] M. I. D'yakonov and V. I. Perel', Zh. Éksp. Teor. Fiz. **58**, 1090 (1970) [Sov. Phys. JETP **31**, 585 (1970)].
[4] I. M. Petrov, Yu. A. Matyugin, S. G. Rautian, and V. P. Chebotarev, Zh. Éksp. Teor. Fiz. **58**, 1243 (1970) [Sov. Phys. JETP **31**, 668 (1970)].

[5] A. D. Akhsakhalyan, S. V. Gaponov, V. I. Luchin, and A. P. Chirinyanov, Zh. Tekh. Fiz. **58**, 1885 (1988) [Sov. Phys. Tech. Phys. **33**, 1146 (1988)].
[6] R. W. Wood, Phys. Z. **10**(13), 425 (1909).
[7] W. Heering, Z. Phys. B: Condens. Matter **20**, 69 (1975).
[8] M. Gröbel and W. Heering, Appl. Phys. **21**, 77 (1980).
[9] C. I. Koester, IEEE J. Quantum Electron. **2**, 580 (1966).
[10] B. B. Boĭko and N. S. Petrov, *Light Reflection from Amplifying and Nonlinear Media* [in Russian], Nauka i Tekhnika, Minsk (1988).
[11] B. A. Veklenko, Zh. Éksp. Teor. Fiz. **96**, 457 (1989) [Sov. Phys. JETP **69**, 258 (1989)].
[12] B. A. Veklenko, R. B. Gusarov, and Yu. B. Sherkunov, Zh. Éksp. Teor. Fiz. **113**, 521 (1998) [JETP **86**, 289 (1998)].
[13] F. T. Arecchi, M. O. Scully, H. Haken, and V. Weidlich, in *Quantum Optics*, R. J. Glauber (Ed.), Academic Press, New York (1969) [Russ. transl., Mir, Moscow (1974)].
[14] B. A. Veklenko, Izv. Vyssh. Uchebn. Zaved. Fiz. **9**, 71 (1983).
[15] M. Born and E. Wolf, *Principles of Optics*, 4th ed., Pergamon Press, Oxford–New York (1970) [Russ. transl., Nauka, Moscow (1973)].
[16] W. Heitler, *The Quantum Theory of Radiation*, 3rd ed., Clarendon Press, Oxford (1954) [Russ. transl., IL, Moscow (1956)].
[17] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics*, Prentice-Hall, Englewood Cliffs, N. J. (1963) [Russ. original, Gostekhizdat, Moscow (1962)].
[18] D. A. Kirzhnits, *Field Theoretical Methods in Many-Body Systems*, Pergamon Press, Oxford (1967) [Russ. original, Gos. Izd. Lit. po Atomnoĭ Nauke i Tekhnike, Moscow (1963)].
[19] B. A. Veklenko, Izv. Vyssh. Uchebn. Zaved. Fiz. **5**, 81 (1978).
[20] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1965)].

Translated by P. Shelnitz

Third-order diamagnetic susceptibilities of hydrogenlike atoms

V. D. Ovsyannikov*) and K. V. Khalyov

Voronezh State University, 394693 Voronezh, Russia
(Submitted 13 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 838–857 (September 1999)

We develop a method for calculating diamagnetic susceptibilities based on higher-order perturbation theory for the wave function and energy of the excited states of the hydrogen atom with degeneracy of arbitrary multiplicity. We derive analytical expressions for third-order matrix elements in the spherical states $|nlm\rangle$ with fixed principal quantum number n and magnetic quantum number m. The formulas for the susceptibilities of doubly degenerate levels are represented in the form of radical-fractional relationships containing polynomials in the principal quantum number. We establish the existence of a monotonic interdependence between the absolute values of susceptibilities of the first three orders. We also present the results of numerical calculations for the states with $n \leqslant 6$ and $m \leqslant 3$ mixed by the field. Finally, for Rydberg states with large n and small m we detect the existence of a discontinuity in the interdependence of the susceptibilities at the boundary between the doublet and equidistant parts of the spectrum of diamagnetic sublevels with opposite parities.
[S1063-7761(99)00709-X]

1. INTRODUCTION

The study of the interaction between atoms and electromagnetic fields is an important avenue of research in atomic physics. The central problem here is the calculation of the Stark and Zeeman effects in the simplest quantum system, the hydrogen atom. The latest achievements in solving this problems are reflected in review articles and monographs (see, e.g., Refs. 1–3). Nevertheless, there are still many problems to be solved. One such problem is the calculation of corrections to the energy of an atom in a magnetic field in higher-order perturbation theory. In contrast to the Stark effect, where all calculations are conveniently done in a parabolic system of coordinates and where the analytical expressions for the higher-order corrections in the form of polynomials in the parabolic quantum numbers of a level were derived more than 20 years ago,[4] so far the results that have been obtained for the Zeeman effect are limited to the first- and second-order perturbations in the diamagnetic interaction[5] and to numerical calculations of higher-order corrections[6,7] or of the exact energy values[8,9] only for specific levels (the ground level and two to three excited levels). The real reason for such a situation is the fact that the complete set of constants of motion for an atom in an electric field can be represented by parabolic quantum numbers, which are also constants of motion for a free atom, while in a magnetic field the set of constants of motion can be found only approximately, to second order in the diamagnetic interaction.[10] Hence calculations of the third- and higher-order corrections for multiply degenerate excited states of hydrogen are fraught with substantial difficulties.

Thus, most of the data on the interaction of an atom and a magnetic field exists in the literature in the form of tables of numerical values of the energies of the hydrogenlike levels in fields with a fixed strength[8,9,11] and cannot be used at such field strengths for other states of atoms or for the same levels but at different field strengths. In view of this, a method that would enable obtaining a closed system of analytical expressions for calculating in a fairly simple manner the shift and splitting of the atomic levels in situations interesting from the practical viewpoint would play a major role.

The special interest in obtaining general formulas for electromagnetic susceptibilities has lately emerged in connection of studies of long-lived highly excited atoms in Rydberg states. The enormous number of such states and the unlimited set of external fields that can act on them makes it impossible to compile tables that would be complete enough to find the energies of these levels and the corresponding frequencies of the spectral lines. Rydberg states with high angular momenta l are almost the same for all atoms, with the result that the formulas derived for hydrogen can be applied to many-electron atoms in highly excited states with magnetic quantum numbers $m > 3$.

In a broad range of magnetic-field strengths encountered in practice, the theory of perturbations in the atom–field interaction proves to be sufficient for calculating the shift of the energy of bound levels (Rydberg levels included). An analysis of the nonlinear Zeeman effect shows that the perturbation-theory series are asymptotic, in which the coefficients, the diamagnetic susceptibilities, alternate in sign and depend strongly on the structure of the unperturbed state. Knowing such susceptibilities makes it possible not only to numerically calculate the energy values but also to determine the limits and errors of such calculations for a fixed value of the field strength. Because the signs of the asymptotic series alternate, the coefficients of the series enable determining the upper and lower limits of the interval inside which the exact value of the energy of the atom in the field lies. The larger the number of the coefficients of the series known, the closer the upper and lower limits are to each other and the more

accurate is the procedure of finding the energy by perturbation-theory techniques. Thus, the importance of calculating higher-order corrections in the perturbation- theory setting is related to the possibility of using these corrections not only to estimate the shift and splitting of the atomic levels but also to monitor the accuracy of other methods of determining the Zeeman energy of atoms.

Substantial progress in calculating the coefficients of perturbation-theory series was achieved by Vaĭnberg *et al.*,[7] who calculated the first 80 diamagnetic susceptibilities for the lower levels of hydrogen (with the principal quantum number $n \leqslant 3$). They used these susceptibilities to build Padé approximants, which enabled them to sum the series for the diamagnetic energy in a range of magnetic-field strengths up to the atomic. However, the fact that their results cannot be applied to higher levels (and this is especially true of Rydberg states) and the significant technical difficulties in using their method for high states only emphasize the need for a method of obtaining general formulas for higher-order susceptibilities in the form of functions of the principal and magnetic quantum numbers.

In the present paper we develop a method for calculating higher-order susceptibilities. The method is based on the successive solution of a system of equations for the wave function and energy of a degenerate state in the corresponding perturbation-theory orders (Sec. 2). In Sec. 3 we derive analytical expressions for the matrix elements of the diamagnetic-interaction operator in the form of polynomials in the principal quantum number n, the orbital quantum number l, and the magnetic quantum number m of a degenerate hydrogenlike state $|nlm\rangle$. The expressions are used to calculate, both analytically and numerically, the third-order magnetic susceptibility $\chi^{(3)}_{nmp\lambda}$. The parameter λ labels the Zeeman states that arise as a result of the mixing, by the diamagnetic interaction, of states with different angular momenta $l=m+p,\ m+p+2...,l_{\max}$ and conserve only the magnetic quantum number m and the parity $P=(-1)^{m+p}$; here $l_{\max}=n-2$ (or $l_{\max}=n-1$) if the parities of l and n are the same (or opposite). For the four lowest components in the set of Zeeman states with a fixed m (nondegenerate states), more precisely, $n=m+1$ of parity $P=(-1)^m$, $n=m+3$ of parity $P=(-1)^{m+1}$, and $n=m+2$ of both parities, the susceptibility is determined only by the diagonal matrix element (the corresponding expressions for $\chi^{(3)}_{nlm}$ can be found in Ref. 12).

In Sec. 4 we give the analytical expressions for third-order diamagnetic susceptibilities and for the eight components that follow on the energy scale with $n=m+3$ and $P=(-1)^m$, $n=m+4$ and $P=\pm 1$, and $n=m+5$ and $P=(-1)^{m+1}$, which are doubly degenerate (in the diamagnetic interaction) states with fixed n, m, and P. The calculation of susceptibilities of states whose degeneracy multiplicity is three or larger is discussed in Sec. 5. The numerical data obtained for states with large n and small m exhibit a discontinuity in the interdependence of the susceptibilities, which develops at the boundary between the doublet and split diamagnetic states of opposite parities, and a difference in the relative rates of their monotonic increase in the two characteristic parts of the diamagnetic spectra.

2. PERTURBATION THEORY FOR THE DEGENERATE LEVELS OF HYDROGEN IN A MAGNETIC FIELD

The main difficulty in calculating higher-order corrections to the energy of hydrogenlike atoms stems from the n^2-fold degeneracy of a state with a given principal quantum number n. The interaction between the atomic moment and a magnetic field splits each state into sublevels with fixed magnetic quantum numbers m. The magnetic quantum number and the parity $P=(-1)^{m+p}$ ($p=0$ or 1) are constants of motion in the magnetic field. Hence the operator of the magnetodipole interaction[1)] $V_m=-\mathbf{m}\cdot\mathbf{B}$, where $\mathbf{m}=-(\mathbf{l}+2\mathbf{s})/2$ is the magnetic moment ($\mathbf{s}$ is the spin moment of the electron), can be incorporated into the unperturbed Hamiltonian, with the result that we can limit ourselves to the subspace of states with fixed n and m and consider only the operator of the diamagnetic interaction

$$V(\mathbf{r})=\frac{(\mathbf{B}\times\mathbf{r})^2}{8}=\frac{B^2}{12}r^2[1-C_{20}(\theta,\varphi)] \tag{1}$$

as the perturbation. Here $C_{20}(\theta,\varphi)=\sqrt{4\pi/5}Y_{20}(\theta,\varphi)$ is a modified spherical function.

The wave function of the a degenerate state of the atom in the field can be written[13]

$$\psi_{nmp}(\mathbf{r})=\sum_{l=m+p}^{l_{\max}} a_l\varphi_{nlm}(\mathbf{r}) - G'_E(\mathbf{r},\mathbf{r}')V(\mathbf{r}')|\psi_{npm}(\mathbf{r}')\rangle, \tag{2}$$

where summation is over all states φ_{nlm} of the same parity from the given nm-shell, and

$$G'_E(\mathbf{r},\mathbf{r}')=G_E(\mathbf{r},\mathbf{r}')-\sum_{l=m+p}^{l_{\max}}\frac{\varphi_{nlm}(\mathbf{r})\varphi^*_{nlm}(\mathbf{r}')}{E_n-E} \tag{3}$$

is the reduced Green's function. The total Green's function $G_E(\mathbf{r},\mathbf{r}')$ is the solution of a Schrödinger equation with a delta-function inhomogeneity,

$$[\hat{H}(\mathbf{r})-E]G_E(\mathbf{r},\mathbf{r}')=\delta(\mathbf{r}-\mathbf{r}'), \tag{4}$$

and can be written as a spectral expansion over the complete set of eigenfunctions of the unperturbed atom (the continuous spectrum included):

$$G_E(\mathbf{r},\mathbf{r}')=\sum_{nlm}\frac{\varphi_{nlm}(\mathbf{r})\varphi^*_{nlm}(\mathbf{r}')}{E_n-E}. \tag{5}$$

Thus, the first term on the right-hand side of Eq. (2), which is a linear combination of all states of a degenerate base with given nmp, is orthogonal to the second term, which allows for the contribution of states from subshells with other principal quantum numbers $n'\neq n$. By iteratings we can represent the wave function (2) in the form of a Brillouin–Wigner series,

$$\psi_{nmp}(\mathbf{r})=\sum_{l=m+p}^{l_{\max}} a_l\sum_{s=0}^{\infty}[-G'_E(\mathbf{r},\mathbf{r}')V(\mathbf{r}')]^s|\varphi_{nlm}(\mathbf{r}')\rangle, \tag{6}$$

where the Green's function depends on the exact energy E. Substituting this series into the Schrödinger equation and

projecting the result on the states of the unperturbed spherical base with fixed principal (n) and magnetic (m) quantum numbers, we arrive at a system of algebraic equations for the coefficients a_l of the linear combination:

$$(E_n-E)a_l+\sum_{l'=m+p}^{l_{\max}} a_{l'}\langle\varphi_{nlm}(\mathbf{r})|\hat{W}(\mathbf{r},\mathbf{r}')|\varphi_{nl'm}(\mathbf{r}')\rangle=0,$$

$$l=m+p,\ldots,l_{\max}, \tag{7}$$

where

$$\hat{W}(\mathbf{r},\mathbf{r}')(E)=V(\mathbf{r})\sum_{s=0}^{\infty}[-G'_E(\mathbf{r},\mathbf{r}')V(\mathbf{r}')]^s$$

$$=V(\mathbf{r})[1+G'_E(\mathbf{r},\mathbf{r}')V(\mathbf{r}')]^{-1} \tag{8}$$

is an integral operator of the atom–field interaction that incorporates all perturbation-theory orders in $V(\mathbf{r})$. To obtain the Rayleigh–Schrödinger series for the energy, we must also expand the Green's function in (6) in a power series in B^2 by using the relationship[13]

$$G'_E(\mathbf{r},\mathbf{r}')=\sum_{N=0}^{\infty}[G'_{E_n}(\mathbf{r},\mathbf{r}')]^{N+1}(E-E_n)^N. \tag{9}$$

By expanding in this manner the operator (8) in powers of $V(\mathbf{r})$, we can find the corresponding expansion for the matrix element $W_{ll'}$ of the system of linear equations (7):

$$W_{ll'}=\langle\varphi_{nlm}(\mathbf{r})|\hat{W}(\mathbf{r},\mathbf{r}')|\varphi_{nl'm}(\mathbf{r}')\rangle.$$

Further calculations of the energy can be done in two ways:[14]

(1) by expanding in a powers of the parameter B^2 the determinant of the secular equation for the system (7),

$$D(E)=\det\|\alpha_{ll'}(E)\|=0, \tag{10}$$

whose elements $\alpha_{ll'}(E)=(E_n-E)\delta_{ll'}+W_{ll'}$ are represented by the series

$$\Delta E=E-E_n=-\sum_{N=1}^{\infty}\frac{\chi_{nm}^{(N)}}{(2N)!}B^{2N},$$

$$W_{ll'}=-\sum_{N=1}^{\infty}\frac{w_{ll'}^{(N)}}{(2N)!}B^{2N}, \tag{11}$$

where $\chi_{nm}^{(N)}$ and $w_{ll'}^{(N)}$ are the diamagnetic susceptibility[15] and the diamagnetic matrix element of the Nth order that are independent of the field strength B; and

(2) by expanding (together with the energy and the matrix element) the coefficients of the linear combination in the system of equations (7) in a power series,

$$a_l=\sum_{N=0}^{\infty}a_l^{(N)}B^{2N}, \tag{12}$$

and by solving the resulting equations for the corrections to the coefficients in each order of B^2, which are then used to determine the energy.

We will employ the second approach, which proves to be easier in practical applications, since the volume of calculations in this approach is proportional to the number of the matrix elements (11), i.e., K^2, irrespective of the perturbation-theory order, while the theory of perturbations for matrices, which is used in the first approach, requires a volume of calculations proportional to the square of this number, i.e., K^4, in the second order, to the cube of this number, i.e., K^6, in the third order, etc., where $K=(l_{\max}-m-p)/2+1$ is the number of dimensions of the state space in which the eigenvalue of the operator (8) is sought. Furthermore, the approach in which the diamagnetic susceptibilities and the expansion coefficients in (12) are found successively also proves to be useful in calculations of the magnetic-field dependent corrections not only for the energy but also for the wave functions. Here the first-order susceptibility $\chi_\lambda^{(1)}$ and the zeroth-order coefficients $a_l^{(0)}(\lambda)$ are the eigenvalues and eigenvectors of the matrix $w_{ll'}^{(1)}$. The discrete parameter $\lambda=1,2,\ldots,K$, which we call the diamagnetic quantum number, labels the states of the atom in the field obtained as a result of mixing and rearrangement of the Ks of the degenerate states of a free atom. The very fact that the quantities in the system of equations (7) can be represented by the series (11) and (12) makes it possible to remove the factor B^2 from (7) in any order N.

The quadratic diamagnetic susceptibility $\chi_\lambda^{(2)}$ appears in the system of equations (7) in the second order in B^2. Substituting (11) and (12) in this system, we arrive at a system of equations for the coefficients $a_l^{(1)}$ that depends on the eigenvalue $\chi_\lambda^{(1)}$ and the corresponding eigenvector composed of the coefficients $a_{l'}^{(0)}(\lambda)$ (both the eigenvalue and the eigenvector was found in the previous stage):

$$\sum_{l'=m+p}^{l_{\max}}a_{l'}^{(1)}(\lambda)(w_{ll'}^{(1)}-\chi_\lambda^{(1)}\delta_{ll'})$$

$$=\sum_{l'=m+p}^{l_{\max}}a_{l'}^{(0)}(\lambda)(\chi_\lambda^{(2)}\delta_{ll'}-w_{ll'}^{(2)}). \tag{13}$$

All the quantities on the right-hand side of (13) are known except for the susceptibility $\chi_\lambda^{(2)}$, which can easily be found by multiplying (13) into $a_l^{(0)*}(\lambda)$. By summing the resulting equations over all l we annihilate the left-hand side, since $\chi_\lambda^{(1)}$ is an eigenvalue corresponding to the eigenvector $\{a_l^{(0)}(\lambda)\}$ of the matrix $w_{ll'}^{(1)}$. Allowing for the completeness condition $\Sigma|a_l^{(0)}(\lambda)|^2=1$ for the basis, we arrive at an expression for the second-order susceptibility:

$$\chi_\lambda^{(2)}=\sum_{l,l'=m+p}^{l_{\max}}a_l^{(0)*}(\lambda)a_{l'}^{(0)}(\lambda)w_{ll'}^{(2)}, \tag{14}$$

which is the second-order diagonal matrix element[16] in the diamagnetic states with a fixed λ and is written in the form of a linear combination of the wave function of the spherical basis, which diagonalizes the diamagnetic Hamiltonian (1).

In the third order in B^2, the coefficients $a_l^{(2)}$ of (12) become the unknown quantities of the system of algebraic equations (7). The equations for $a_l^{(2)}$ are similar to (13). Using the same method as we did in deriving Eq. (14), we annihilate the left-hand side of this system and arrive at an expression for the third-order susceptibility:[14]

$$\chi_\lambda^{(3)}=\sum_{l,l'=m+p}^{l_{\max}}[a_l^{(0)*}(\lambda)a_{l'}^{(0)}(\lambda)w_{ll'}^{(3)}(\lambda)+a_l^{(0)*}(\lambda)a_{l'}^{(1)}(\lambda)w_{ll'}^{(2)}]. \tag{15}$$

The first term on the right-hand side contains a quadratic combination of the coefficients $a_l^{(0)}$ and the matrix elements $w_{ll'}^{(3)}(\lambda)$ (which, in contrast to $w_{ll'}^{(1)}$ and $w_{ll'}^{(2)}$, depend on the susceptibility $\chi_\lambda^{(1)}$). The second term is a bilinear form of the coefficients $a_l^{(0)}$ and $a_l^{(1)}$ of the first and second order, respectively; the coefficients of this bilinear form are the matrix elements $w_{ll'}^{(2)}$.

Thus, the transition from second to third order requires two operators:

(1) the solution of the system of equations (7) in the second order in B^2, which leads to the inhomogeneous equation (13) for $a_l^{(1)}(\lambda)$; and

(2) the calculation of third-order matrix elements in the diamagnetic interaction, $w_{ll'}^{(3)}(\lambda)$, which contain a contribution from the corrections to the second-order matrix elements due to the expansion (9) of the Green's function.

3. DIAMAGNETIC MATRIX ELEMENTS OF A HYDROGENLIKE ATOM

Calculating the diamagnetic interaction energy in degenerate states involves diagonalizing the matrix of the operator (1) in the states of the shell of the hydrogenlike atom with given principal magnetic number n and magnetic quantum number m. In first order such diagonalization has been carried out both in the spherical basis of the states $|nlm\rangle$ and in the basis with fixed parabolic quantum numbers (see, e.g., Ref. 3). In the spherical basis, the expressions for the first-order matrix elements are

$$w_{ll}^{(1)}=-\left(\frac{n}{2Z}\right)^2\frac{l^2+l-1+m^2}{(2l-1)(2l+3)}[5n^2+1-3l(l+1)], \tag{16}$$

$$w_{ll+2}^{(1)}=w_{l+2,l}^{(1)}=\frac{5}{2}\left(\frac{n}{2Z}\right)^2\times\left[\frac{(l+1-m)_2(l+1+m)_2(n-l-2)_2(n+l+1)_2}{(2l+1)(2l+3)^2(2l+5)}\right]^{1/2}. \tag{17}$$

Here we have used the standard Pochhammer symbol:[17] $(c)_n=c(c+1)\cdots(c+n-1)$.

Delande and Gay[18] were the first to calculate the second-order matrix elements. They used an effective Hamiltonian (for which they wrote an explicit formula) that took into account the additional symmetry of hydrogenlike states in a magnetic field. Two years later Grozdanov and Taylor[5] diagonalized this Hamiltonian in a basis of states with fixed parabolic quantum numbers. The corrections to the energy for nondegenerate states were found to coincide with the corresponding diagonal matrix elements. For doubly degenerate states, the expression for the diamagnetic energy can also be found analytically by using the roots of the corresponding quadratic equation.

The calculation of second-order diamagnetic matrix elements has also been done via a Sturm expansion of the reduced Green's function in the basis of states with fixed moments[14,15] without using an effective Hamiltonian. By separating the orientational dependence in the matrix elements $w_{ll'}^{(2)}$ via integration over the angular variables we can conveniently express these elements in terms of their irreducible parts as follows:

$$w_{ll}^{(2)}=\beta_{nl}^{(0)}+\frac{3m^2-l(l+1)}{l(2l-1)}\beta_{nl}^{(2)}+\frac{3(l^2+2l-5m^2)(l^2-5m^2-1)-10m^2(4m^2-1)}{l(2l-1)(2l-2)(2l-3)}\times\beta_{nl}^{(4)}, \tag{18}$$

$$w_{ll+2}^{(2)}=-\sqrt{\frac{(l+1-m)_2(l+1+m)_2(n-l-2)_2(n+l+1)_2}{(2l-1)^2(2l+1)(2l+3)^2(2l+5)(2l+7)^2}}\times\frac{n^6}{32Z^6}[\gamma_{nl}^{(0)}+m^2\gamma_{nl}^{(2)}], \tag{19}$$

$$w_{ll+4}^{(2)}=\sqrt{\frac{(l+1-m)_4(l+1+m)_4(n-l-4)_4(n+l+1)_4}{(2l+1)(2l+3)^2(2l+5)^2(2l+7)^2(2l+9)}}\times\frac{205n^6}{128Z^6}. \tag{20}$$

The irreducible parts β_{nl} and γ_{nl} can be expressed in terms of the second-order radial matrix elements of the operator r^2, i.e., $\langle nl|r^2g_{ll'}^{(n)}(r,r')r'^2|nl'\rangle$, which can be calculated analytically by using a Sturm expansion of the Coulomb Green's function:[12,14]

$$g_l^{(n)}(r,r')=\frac{4Z}{n}\left\{\sum_{k\neq n_r}^{\infty}\frac{k!}{(k+2l+1)!}\frac{f_{kl}\left(\frac{2Zr}{n}\right)f_{kl}\left(\frac{2Zr'}{n}\right)}{k+l+1-n}+\frac{n_r!}{(N=l)!n}+\left[\frac{5}{2}f_{n_rl}\left(\frac{2Zr}{n}\right)f_{n_rl}\left(\frac{2Zr'}{n}\right)+r\frac{df_{n_rl}\left(\frac{2Zr}{n}\right)}{dr}f_{n_rl}\left(\frac{2Zr'}{n}\right)+f_{n_rl}\left(\frac{2Zr}{n}\right)r'\frac{df_{n_rl}\left(\frac{2Zr}{n}\right)}{dr'}\right]\right\}. \tag{21}$$

Due to the orthogonality of the Sturm function

$$f_{kl}(x)=e^{-x/2}x^lL_k^{2l+1}(x) \tag{22}$$

and the radial wave function

$$R_{nl}(r)=\frac{2Z^{3/2}}{n^2}\sqrt{\frac{n_r!}{(n+l)!}}f_{n_rl}\left(\frac{2Zr}{n}\right), \tag{23}$$

which follows from the orthogonality of the Laguerre polynomials $L_k^\alpha(x)$ (see Ref. 17), the Sturm series (21) is truncated. In particular,

$$\left\langle f_{kl}\left(\frac{2Zr}{n}\right)\middle| r^2 \middle| f_{k'l}\left(\frac{2Zr}{n}\right)\right\rangle$$
$$=\left(\frac{n}{2Z}\right)^5\int_0^\infty e^{-x}x^{2l+4}L_k^{2l+1}(x)L_{k'}^{2l+1}(x)dx$$
$$=\left(\frac{n}{2Z}\right)^5\frac{(k+2l+1)!}{k!}\{-(k-2)_3\delta_{k'k-3}+6(k+l)$$
$$\times(k-1)_2\delta_{k'k-2}-3k[5k(k+2l+1)+4l(l+1)$$
$$+2]\delta_{k'k-1}+[(k+2l+2)_3+9k(k+2l+2)_2+9(k$$
$$-1)_2(k+2l+2)+(k-2)_3]\delta_{k'k}-3(k+2l+2)$$
$$\times[5k(k+2l+3)+4l^2+14l+12]\delta_{k'k+1}+6(k+l$$
$$+2)(k+2l+2)_2\delta_{k'k+2}-(k+2l+2)_3\delta_{k'k+3}\}. \quad (24)$$

where $\delta_{k'k}$ is the Kronecker delta.

Thus, the irreducible parts in (18)–(20) can be represented by polynomials of the principal and orbital quantum numbers:[14]

$$\beta_{nl}^{(0)}=\frac{n^6}{240Z^6}\{5n^2[97n^2-33l(l+1)+365]-12(21l^4$$
$$+42l^3+179l^2+158l+60)\}, \quad (25)$$

$$\beta_{nl}^{(2)}=\frac{ln^6}{336Z^6(2l+3)}\{n^2[802n^2-1005l(l+1)+2000]$$
$$+3(175l^4+350l^3+523l^2+348l-480)\}, \quad (26)$$

$$\beta_{nl}^{(4)}=\frac{l(l-1)n^6}{1120Z^6(2l+3)(2l+5)}\{15n^2[41n^2+22l(l+1)$$
$$-185]-1001l^3(l+2)+3421l^2+4422l-2160\}, \quad (27)$$

$$\gamma_{nl}^{(0)}=n^2[429l(l+3)-802]+237l^3(l+6)+3145l^2$$
$$+3036l-2484, \quad (28)$$

$$\gamma_{nl}^{(2)}=5(41n^2+37l^2+111l-75). \quad (29)$$

These expressions simplify substantially when l and m take specific numerical values (or are expressed in terms of n). The solution of the system of equations (7) also simplifies if K is moderate. The corresponding analytical expressions for nondegenerate ($K=1$) and doubly degenerate ($K=2$) states can be found in Refs. 14 and 15. Note that using the spherical basis has an important advantage over using the parabolic basis, since the operator (1) is of even parity and hence the wave functions $|nlm\rangle$, which have a well-defined parity, automatically take into account the symmetry of the atom in a magnetic field. This feature of diamagnetic states was probably the reason why Grozdanov and Taylor,[5] who used an effective Hamiltonian in the parabolic basis, examined only three out of the four sets of doubly degenerate states of the n-shell.

When calculating the third-order energy, one must take into account the corrections to the expansion coefficients in (12) that determine their dependence on the amplitude B of the magnetic field. To find $a_l^{(1)}(\lambda)$ it is enough to solve the system of equation (13) depending on the matrices $w_{ll'}^{(1)}$ and $w_{ll'}^{(2)}$. Since the rank of the matrix consisting of the coefficients of the left-hand side of the system (13) is one unit smaller than the number of unknowns, to solve the system we must use the normalization condition for the $a_l^{(1)}(\lambda)$:

$$\sum_{l=m+p}^{l_{\max}} a_l^{(0)*}(\lambda)a_l^{(1)}(\lambda)=0.$$

What sets the matrix $w_{ll'}^{(3)}$ apart from the first- and second-order matrices is its dependence on $\chi_\lambda^{(1)}$, the eigenvalue of the matrix $w_{ll'}^{(1)}$. After calculating the integrals over the angular variables, we see that the matrix elements $w_{ll'}^{(3)}$ can be expressed in terms of linear combinations of the radial matrix elements $p_{l;l_1,l_2;l'}^{2q2}=\langle nl|r^2g_{l_1}^{(n)}r^qg_{l_2}^{(n)}r^2|nl'\rangle$ and the first-order susceptibility $\chi_\lambda^{(1)}$:

$$w_{ll}^{(3)}(\lambda)=-\frac{45}{4}\{\mathscr{Q}_{l-2,m}\mathscr{P}_{l-2,m}^2p_{l;l-2,l-2;l}^{222}$$
$$+2\mathscr{Q}_{l,m}\mathscr{P}_{l-2,m}^2p_{l;l-2,l;l}^{222}+\mathscr{Q}_{l,m}^3p_{l;l,l;l}^{222}$$
$$+2\mathscr{Q}_{l,m}\mathscr{P}_{l,m}^2p_{l;l+2,l;l}^{222}+\mathscr{Q}_{l+2,m}\mathscr{P}_{l,m}^2p_{l;l+2,l+2;l}^{222}$$
$$+2\chi_\lambda^{(1)}[\mathscr{P}_{l-2,m}^2p_{l;l-2,l-2;l}^{202}+\mathscr{Q}_{l,m}^2p_{l;l,l;l}^{202}$$
$$+\mathscr{P}_{l,m}^2p_{l;l+2,l+2;l}^{202}]\}, \quad (30)$$

$$w_{ll+2}^{(3)}(\lambda)=-\frac{45}{4}\mathscr{P}_{l,m}\{\mathscr{P}_{l-2,m}^2p_{l;l-2,l;l+2}^{222}+\mathscr{Q}_{l,m}^2p_{l;l,l;l+2}^{222}$$
$$+\mathscr{Q}_{l,m}\mathscr{Q}_{l+2,m}p_{l;l,l+2;l+2}^{222}+\mathscr{P}_{l,m}^2p_{l;l+2,l;l+2}^{222}$$
$$+\mathscr{Q}_{l+2,m}^2p_{l;l+2,l+2;l+2}^{222}$$
$$+\mathscr{P}_{l+2,m}^2p_{l;l+2,l+4;l+2}^{222}+2\chi_\lambda^{(1)}[\mathscr{Q}_{l,m}p_{l;l,l;l+2}^{202}$$
$$+\mathscr{Q}_{l+2,m}p_{l;l+2,l+2;l+2}^{202}]\}, \quad (31)$$

$$w_{ll+4}^{(3)}(\lambda)=-\frac{45}{4}\mathscr{P}_{l,m}\mathscr{P}_{l+2,m}$$
$$\times\{\mathscr{Q}_{l+2,m}p_{l;l+2,l+2;l+4}^{222}\mathscr{Q}_{l,m}p_{l;l,l+2;l+4}^{222}$$
$$+\mathscr{Q}_{l+4,m}p_{l;l+2,l+4;l+4}^{222}$$
$$+2\chi_\lambda^{(1)}p_{l;l+2,l+2;l+4}^{202}\}, \quad (32)$$

$$w_{ll+6}^{(3)}(\lambda)=-\frac{45}{4}\mathscr{P}_{l,m}\mathscr{P}_{l+2,m}\mathscr{P}_{l+4,m}p_{l;l+2,l+4;l+6}^{222}$$
$$=\frac{25n^{10}(902n^2+88l^2+616l+7125)}{2^{17}Z^{10}(l+3/2)_5}$$
$$\times\sqrt{\frac{(n-l-6)_6(n+l+1)_6(l+1-m)_6(l+1+m)_6}{(2l+1)(2l+13)}}. \quad (33)$$

In these expressions, the factors $\mathscr{Q}_{l,m}$ and $\mathscr{P}_{l,m}$ are integrals over the angular variables:

$$\mathscr{Q}_{l,m}=\frac{1}{3}\langle lm|1-C_{20}(\theta,\varphi)|lm\rangle=\frac{l^2+l+m^2-1}{(2l-1)(2l+3)},$$

TABLE I. Coefficients of the polynomials $P^{(q)}_{\nu_1,\nu_2,k}=\Sigma c_{ss'}n^s l^{s'}$ that determine the diagonal ($k=0$) radial matrix elements $p^{2q2}_{l;l+\nu_1,l+\nu_2;l}$ according to (35).

$c_{ss'}$	$P^{(2)}_{-2,-2,0}$	$P^{(2)}_{-2,0,0}$	$P^{(2)}_{0,0,0}$	$P^{(2)}_{0,2,0}$	$P^{(2)}_{2,2,0}$	$P^{(0)}_{-2,-2,0}$	$P^{(0)}_{0,0,0}$	$P^{(0)}_{2,2,0}$
c_{80}	4510	4510	4510	4510	4510	0	0	0
c_{62}	7346	−5066	−6066	−5066	7346	0	0	0
c_{61}	−126818	−60736	−6066	50604	141510	0	0	0
c_{60}	261905	231967	204882	287637	396069	1804	1804	1804
c_{44}	−13958	−3654	966	−3654	−13958	0	0	0
c_{43}	108832	95478	1932⁻	−110094	−164664	0	0	0
c_{42}	113737	−270404	−200658	−578762	−296507	4424	−1344	4424
c_{41}	−1405413	−608068	−201624	−233790	1250559	−39368	−1344	48216
c_{40}	1455860	1154248	944994	1392780	2852220	55098	37528	98890
c_{26}	1150	2930	530	2930	1550	0	0	0
c_{25}	28110	−36300	1590	53880	−18810	0	0	0
c_{24}	−428465	80495	22070	305945	−545765	−4340	−420	−4340
c_{23}	1409120	557700	41490	185880	−3373080	16600	−840	−33960
c_{22}	−809656	−1025,439	−594154	−1808619	−7865656	44760	−20700	−31080
c_{21}	−1563937	−766264	−614634	−2436654	−6127845	−181100	−20280	203460
c_{20}	1039645	771363	654758	74213	−70219	111626	49516	316546
c_{08}	840	1280	60	1280	840	0	0	0
c_{07}	−14940	−1130	240	11370	21660	0	0	0
c_{06}	87055	−12490	3802	31260	215155	−1600	−40	−1600
c_{05}	−138135	−86480	10566	106950	1021245	29920	−120	−29520
c_{04}	−329135	215159	34108	589359	2249068	−88570	−4012	−212170
c_{03}	1107453	262372	50886	1324494	1268409	158340	−7824	−743820
c_{02}	−651797	−345051	183410	895807	−2924513	−87274	−19420	−1316914
c_{01}	−175392	−190332	−207132	−50640	−4442784	−12912	−15528	−1100136
c_{00}	31680	33912	36360	−91620	−1640340	2592	3024	−340200

$$\mathscr{P}_{l,m}=\frac{1}{3}\langle lm|1-C_{20}(\theta,\varphi)|l+2m\rangle$$

$$=-\sqrt{\frac{[(l+1)^2-m^2][(l+2)^2-m^2]}{4(2l+1)(2l+3)^2(2l+5)}}. \tag{34}$$

Only the matrix elements $w^{(3)}_{ll'}(\lambda)$ with $l'=l\pm 6$, which are farthest from the diagonal of the matrix, are independent of $\chi^{(1)}_\lambda$ and can be written in the closed form (33). The matrix elements $w^{(3)}_{ll'}(\lambda)$ with $|l'-l|\leqslant 4$ depend on $\chi^{(1)}_\lambda$ and are written in (30)–(32) in the form of linear combinations of the radial matrix elements $p^{2q2}_{l;l_1,l_2;l'}$, where $q=0$ or 2. These quantities, as well as the second-order matrix elements, can be calculated analytically via the Sturm expansion (21) of the Green's function and the property of orthogonality of the Sturm functions [see Eq. (24)].

We note the following features in the dependence of these quantities on the principal and orbital quantum numbers that manifest themselves in the above analytical expressions:

(1) All matrix elements contain the same common factor n^{8+q}.

(2) The seven diagonal elements $p^{222}_{l;l_1,l_2;l}$ (four of these are pairwise equal) and the three elements $p^{202}_{l;l_1,l_1;l}$ do not contain radicals. The off-diagonal elements $p^{222}_{l;l_1,l_2;l+k}$ (six with $k=2$, three with $k=4$, and one with $k=6$) and $p^{202}_{l;l_1,l_1;l+k}$ (two with $k=2$ and one with $k=4$) contain radical factors of the product of two Pochhammer symbols with the index equal to k, similar to the corresponding factors in (17), (19), (20), and (33).

(3) The radial elements obey the symmetry relationship $p^{2q2}_{l;l_1,l_2;l'}=p^{2q2}_{l';l_2,l_1;l}$, which corresponds to the symmetry of the matrix, $w^{(3)}_{ll'}(\lambda)=w^{(3)}_{l'l}(\lambda)$. This makes it possible to represent all finite elements by the four general expressions (30)–(33).

(4) For all matrix elements the asymptotic ($n\gg 1$ and $n\gg l$) dependence is the same: $p^{222}_{l;l_1,l_2;l'}\sim 2255n^{18}/576$ and $p^{202}_{l;l_1,l_1;l'}\sim 451n^{14}/288$.

These properties make it possible to write a general expression for the matrix elements in (30)–(33):

$$p^{2q2}_{l;l+\nu_1,l+\nu_2;l+k}=\frac{n^{8+q}}{1152Z^{8+q}}\sqrt{(n-l-k)_k(n+l+1)_k}$$

$$\times P^{(q)}_{\nu_1,\nu_2,k}(n,l), \tag{35}$$

where the $P^{(q)}_{\nu_1,\nu_2,k}(n,l)$ are $(6+q-k)$-degree polynomials in the principal and orbital quantum numbers:

$$P^{(q)}_{\nu_1\nu_2,k}(n,l)=\sum_{s=0}^{S}\sum_{s'=0}^{S'}c_{ss'}n^s l^{s'}, \tag{36}$$

with s taking only even values from zero to $S=6+q-k$, and s' running through all integral values (even and odd) from zero to $S'=6+q-k-s$. Thus, the number of terms (and hence the number of the coefficients $c_{ss'}$) in the polynomial is $[4+(q-k)/2]^2$. The numerical values of the coefficients $c_{ss'}$ are listed in Tables I–III. Polynomials are grouped in the tables by the values of k in accordance with the expressions (30)–(32) for the matrix elements $w^{(3)}_{ll+k}$ where they appear. Table I lists 25 coefficients for each of

TABLE II. Coefficients of the polynomials $P^{(q)}_{\nu_1,\nu_2,k}$ that determine the off-diagonal ($k=2$) radial matrix elements $p^{2q2}_{l;l+\nu_1,l+\nu_2;l+2}$ according to (35).

$c_{ss'}$	$P^{(2)}_{-2,0,2}$	$P^{(2)}_{0,0,2}$	$P^{(2)}_{0,2,2}$	$P^{(2)}_{2,0,2}$	$P^{(2)}_{2,2,2}$	$P^{(2)}_{2,4,2}$	$P^{(0)}_{0,0,2}$	$P^{(0)}_{2,2,2}$
c_{60}	4510	4510	4510	4510	4510	4510	0	0
c_{42}	9680	930	−3732	−5820	930	9680	0	0
c_{41}	−97800	−50880	−11196	−17460	56460	155880	0	0
c_{40}	144320	1230047.5	193063	188365	284057.5	524840	1804	1804
c_{24}	−8930	−1270	−1650	−210	−1270	−8930	0	0
c_{23}	36020	7068	−9900	−1260	−22308	−143180	0	0
c_{22}	128820	49369	−100140	−252060	−82823	−677580	1380	1380
c_{21}	−504230	−338505	−255870	−750510	306723	−659830	−11988	20268
c_{20}	315990	256130.5	661135	282865	1422260.5	1292190	15884	64268
c_{06}	−940	−210	−280	−5680	−210	−940	0	0
c_{05}	21340	924	−2520	−51120	−4704	−38260	0	0
c_{04}	−120620	−2658.5	−8395	−80665	−44868.5	−567620	200	200
c_{03}	209780	8697	−12570	282810	−237159	−4085420	−2064	4464
c_{02}	−12210	80199.5	−145543	617135	−642262.5	−15317610	5320	34696
c_{01}	−156510	−170328	−414084	111360	−550812	−28620990	−13560	122808
c_{00}	30780	332976	107172	−99540	437976	−20914740	3024	163512

the $P^{(2)}_{\nu_1,\nu_2,0}(n,l)$ polynomials and 16 coefficients for each of the $P^{(0)}_{\nu_1,\nu_2,0}(n,l)$ polynomials, Table II lists 16 coefficients for each of the $P^{(2)}_{\nu_1,\nu_2,2}(n,l)$ polynomials and 9 for each of the $P^{(0)}_{\nu_1,\nu_2,2}(n,l)$ polynomials, and Table III lists 9 coefficients for each of the $P^{(2)}_{\nu_1,\nu_2,4}(n,l)$ polynomials and 4 for each of the $P^{(0)}_{2,2,4}(n,l)$ polynomials. The polynomial $P^{2}_{2,4,6}(n,l)$ is written explicitly in (33).

The calculations of the matrix elements (35) and the polynomial factors (36) representing these elements were done by standard procedures used in transforming polynomial expressions (factorization and collecting like terms), which can employed through the use of computer programs such as Maple or Mathematica.

4. SUSCEPTIBILITIES OF DOUBLY DEGENERATE STATES

For the states $|nmp\rangle$ with $n=m+1$ and $p=0$, $n=m+2$ and $p=0,1$, and $n=m+3$ and $p=1$ (''nondegenerate'' states), the matrix $w^{(N)}_{ll'}$ consists of one element, which determines the corresponding susceptibility. Thus, by substituting the corresponding values of l and m in (16), (18), and (30) we have working analytical expressions for $\chi^{(N)}$ of these states up to the third order inclusive, formulas that coincide with those known from the literature.[3,5,12,15]

TABLE III. Coefficients of the polynomials $P^{(q)}_{\nu_1,\nu_2,k}$ that determine the off-diagonal ($k=4$) radial matrix elements $p^{2q2}_{l;l+\nu_1,l+\nu_2;l+4}$ according to (35).

$c_{ss'}$	$P^{(2)}_{0,2,4}$	$P^{(2)}_{2,2,4}$	$P^{(2)}_{2,4,4}$	$P^{(0)}_{2,2,4}$
c_{40}	4510	4510	4510	0
c_{22}	2574	−514	2574	0
c_{21}	−31476	−2570	57216	0
c_{20}	45962	96269	267692	1804
c_{04}	116	504	116	0
c_{03}	−4334	5040	6654	0
c_{02}	15493	−3299	97903	176
c_{01}	−43015	−79495	580995	880
c_{00}	10170	27405	1226820	11610

For states with $n=m+3$ and $p=0$, $n=m+4$ and $p=0,1$, and $n=m+5$ and $p=1$ (doubly degenerate states), for which the diamagnetic matrix is 2-by-2, we can also derive analytical formulas for the susceptibility in a form that contains square roots of polynomials constructed from the matrix elements (16) and (17), (18) and (19), and (30) and (31). All such expressions for $\chi^{(N)}$ can represented in the general form

$$\chi^{(1)}_{nmp\pm}=-\frac{n^2}{4Z^2}\left[Q^{(m,p)}_2(n)\pm\sqrt{R^{(m,p)}_2(n)}\right], \tag{37}$$

$$\chi^{(2)}_{nmp\pm}=\frac{n^6}{16Z^6}\left[Q^{(m,p)}_4(n)\pm\frac{R^{(m,p)}_4(n)}{\sqrt{R^{(m,p)}_2(n)}}\right], \tag{38}$$

$$\chi^{(3)}_{nmp\pm}=-\frac{5n^{10}}{64Z^{10}}\left[Q^{(m,p)}_6(n)\pm\frac{R^{(m,p)}_8(n)}{[R^{(m,p)}_2(n)]^{3/2}}\right], \tag{39}$$

where $Q^{(m,p)}_k(n)$ and $R^{(m,p)}_k(n)$ are k-degree polynomials in the principal quantum number n. The explicit expressions for the polynomials of the four doublet states determined by specific sets of the quantum numbers m and p are listed in Table IV. The coefficients of the leading powers of the polynomials are the same for all four sets and increases with the degree of the polynomial. We also note the same order in which the signs alternate: the two leading powers in Q are positive and the signs alternate for the lower powers down; for R the signs of any two neighboring terms are opposite.

The formulas for the $\chi^{(1)}$ and $\chi^{(2)}$ constructed from the polynomials $Q^{(m,p)}_2(n)$, $R^{(m,p)}_2(n)$, $Q^{(m,p)}_4(n)$, and $R^{(m,p)}_4(n)$ coincide with those known from the literature.[5,14]

The explicit expressions for the polynomials $Q^{(m,p)}_6(n)$ and $R^{(m,p)}_8(n)$ reveal the general properties of third-order diamagnetic energy. In particular, the dependence of $\chi^{(3)}$ of doubly degenerate levels on the principal quantum number n has the same asymptotic structure as that of nondegenerate states,[12] so that for $n\gg 1$ the leading term of the asymptotic expansion,

$$\Delta E^{(3)}\approx\frac{3}{128}n^{16}B^6, \tag{40}$$

TABLE IV. Polynomials $Q_k^{(m,p)}(n)$ and $R_k^{(m,p)}(n)$ that determine the diamagnetic susceptibilities (37)–(39) of doubly degenerate states with fixed n, m, and p.

m,p	Polynomial
$n-3,0$	$Q_2=n^2+3n-7$ $R_2=16n^2-48n+41$ $Q_4=12n^4+117n^3-315n^2+597n-714$ $R_4=480n^4-1776n^3+3948n^2-6477n+4734$ $Q_6=216n^6+3915n^5-8631n^4+32871n^3-69750n^2+76812n-65736$ $R_8=237312n^8-1547904n^7+6033408n^6-18774360n^5+43218888n^4-68241357n^3+72834702n^2-4997076n+17036568$
$n-4,1$	$Q_2=n^2+2n-11$ $R_2=16n^2-88n+136$ $Q_4=12n^4+96n^3-488n^2+1192n-2100$ $R_4=480n^4-2968n^3+8872n^2-19960n+23880$ $Q_6=216n^6+3366n^5-17408n^4+61238n^4177212n^2+302016n-350064$ $R_8=237312n^8-295417n^7+17745808n^6-73238368n^5+234785744n^4-567694784n^3+959084336n^2-1034609184n+557701488$
$n-4,0$	$Q_2=n^2+6n-19$ $R_2=16n^2-40n+40$ $Q_4=12n^4+216n^3-840n^2+1920n-3180$ $R_4=480n^4-1768n^3+4360n^2-7624n+7656$ $Q_6=216n^6+7074n^5-27408n^4+102306n^3-283860n^2+435072n-476684$ $R_8=237312n^8-1173568n^7+3714448n^6-11527840n^5+31965008n^4-65148608n^3+93301808n^2-84091488n+39910896$
$n-5,1$	$Q_2=n^2+5n-27$ $R_2=16n^2-80n+145$ $Q_4=12n^4+195n^3-1147n^2+3295n-7182$ $R_4=480n^4-2960n^3+10444n^2-28835n+43578$ $Q_6=216n^6+6525n^5-43303n^4+177725n^3-600754n^2+1257900n-1789128$ $R_8=237312n^8-2578840n^7+14262784n^6-61089320n^5+243111464n^4-794683915n^3+1801819010n^2-2490302340n+1689045624$

proves to be the same for all sublevels of a hydrogenlike shell with fixed principal quantum number and projection of angular momentum, from $m=n-5$ to $m=n-1$. The overall shift is accompanied by a splitting into sublevels, determined by the terms in polynomial (39) that follow the leading term.

5. RESULTS OF NUMERICAL CALCULATIONS AND A DISCUSSION

The analytical expressions for the susceptibilities of nondegenerate and doubly degenerate diamagnetic state with a fixed n make it possible to establish the asymptotic behavior (for $n\gg 1$) common to all the sublevels:

$$\chi^{(1)}_{nm\lambda}\sim -\frac{n^4}{4Z^2},\quad \chi^{(2)}_{nm\lambda}\sim \frac{3n^{10}}{4Z^6},\quad \chi^{(3)}_{nm\lambda}\sim -\frac{135n^{16}}{8Z^{10}}. \tag{41}$$

This yields a relationship between the binding energy E_n and the diamagnetic corrections of the first three orders:

$$|E_n|:\Delta E^{(1)}_{nm\lambda}:|\Delta E^{(2)}_{nm\lambda}|:\Delta E^{(3)}_{nm\lambda} \approx 1:\left(\frac{n^3B}{2Z^2}\right)^2:\left(\frac{n^3B}{2Z^2}\right)^4:3\left(\frac{n^3B}{2Z^2}\right)^6. \tag{42}$$

The third-order correction in this relationship makes obvious, in particular, the asymptotic nature of the perturbation-theory series for the diamagnetic energy: the asymptotic factor for this correction is three times larger than the similar factor for the second order correction, which in turn coincides with the factor for the first-order correction.

Due to this asymptotic nature, the last term in the series that is taken into account determines the accuracy with which perturbation theory approximates the true energy value $E^{\text{exact}}_{nm\lambda p}(B)$ and makes it possible to determine the range of applicability of the corrections of the previous orders. The alternation in the signs of the series allows us to find the upper and lower bounds on the exact energy value, which meet the condition

$$E^{(2N)}_{nm\lambda p}(B)<E^{\text{exact}}_{nm\lambda p}(B)<E^{(2N+1)}_{nm\lambda p}(B), \tag{43}$$

where

$$E^{(s)}_{nm\lambda p}(B)=E_n-\sum_{k=1}^{s}\frac{\chi^{(k)}_{nm\lambda p}}{2k!}B^{2k} \tag{44}$$

is the energy value obtained in sth-order perturbation theory. The binding energy E_n incorporates the self-energy of the free atom and the paramagnetic energy $E_{\text{par}}=B(m+2m_s)/2$.

Table V lists the numerical values of the diamagnetic susceptibilities of the degenerate hydrogen states with $n\leqslant 6$. The values of the parameter λ are selected in accordance with the absolute values of the susceptibilities $\chi^{(1)}$: the minimum value of $|\chi^{(1)}|$ corresponds to $\lambda=1$, and as we move to each next state in $\chi^{(1)}$ the value of λ increases by one unit, so that $|\chi^{(1)}_1|<|\chi^{(1)}_2|<\ldots<|\chi^{(1)}_K|$. The data in Table V together with the data on the nondegenerate sublevels of the same n-shells constitute the complete set of data on the Zeeman effect for the first six levels of a hydrogenlike atom to within corrections of order B^6 inclusive.

Note the rigorous correlation between the absolute values of the susceptibilities of the diamagnetic sublevels: the larger values of $\chi^{(2)}$ and $\chi^{(3)}$ correspond to states with larger values of $\chi^{(1)}$, i.e., for $\lambda>\lambda'$ we have $|\chi^{(N)}_{\lambda}|>|\chi^{(N)}_{\lambda'}|$ for $N=1,2,3$. The maximum value of the susceptibility for a given n corresponds to one of the states with $m=0$ (in Table V these values are printed in boldface type for $n=4,5,6$). Here the difference between the maximum and minimum absolute values of the susceptibility for states with a given n increases with n and with the perturbation-theory order N and follows the approximate relationship $|\chi^{(N)}_{nmpK}/\chi^{(N)}_{nmp1}|\sim 2^{n-1}(n-m-p)$.

The data in Table V can be used to find the limits of applicability (in the magnetic-field strength) of specific perturbation-theory orders in calculations of the energy of an atom in a magnetic field. It occurs that the limit of applicability for first-order diamagnetic corrections, B_1, is approximately twice the limit for the second order, B_2, i.e., $B_1>B_2$, as it should be for an asymptotic series.

TABLE V. Diamagnetic susceptibilities of the degenerate states of hydrogen with $n \leqslant 6$.

$nmp\lambda$	$-\chi^{(1)}_{nmp\lambda}$	$\chi^{(2)}_{nmp\lambda}$	$-\chi^{(3)}_{nmp\lambda}$
3001	1.034(1)	2.442(4)	5.315(8)
3002	3.916(1)	1.918(5)	7.165(9)
4011	2.670(1)	2.776(5)	2.602(10)
4012	7.730(1)	1.809(6)	2.717(11)
4001	3.736(1)	8.116(5)	1.035(11)
4002	**1.306(2)**	**3.299(6)**	**6.143(11)**
4101	4.301(1)	5.623(5)	6.487(10)
4102	1.250(2)	3.121(6)	5.463(11)
5001	5.668(1)	2.141(6)	6.368(11)
5002	1.375(2)	1.023(7)	4.863(12)
5003	**3.308(2)**	**3.045(7)**	**2.035(13)**
5011	6.849(1)	4.860(6)	2.041(12)
5012	2.190(2)	1.954(7)	1.077(13)
5111	8.876(1)	3.435(6)	1.220(12)
5112	2.112(2)	1.836(7)	1.007(13)
5101	1.282(2)	9.593(6)	4.455(12)
5102	3.218(2)	2.945(7)	1.956(13)
5201	1.176(2)	5.867(6)	2.494(12)
5202	2.948(2)	2.640(7)	1.719(13)
6011	1.043(2)	1.254(7)	1.012(13)
6012	2.300(2)	4.136(7)	5.240(13)
6013	5.027(2)	1.322(8)	2.153(4)
6001	1.152(2)	2.168(7)	2.463(13)
6002	3.420(2)	8.350(7)	1.172(14)
6003	**7.038(2)**	**1.884(8)**	**3.638(14)**
6101	1.643(2)	1.614(7)	1.439(13)
6102	3.330(2)	7.787(7)	1.090(14)
6103	6.907(2)	1.843(8)	3.541(14)
6111	2.113(2)	4.141(7)	5.047(13)
6112	4.907(2)	1.275(8)	2.067(14)
6211	2.109(2)	2.536(7)	2.709(13)
6212	4.551(2)	1.133(8)	1.808(14)
6201	3.025(2)	6.375(7)	8.665(13)
6202	6.615(2)	1.717(8)	3.251(14)
6301	2.598(2)	3.864(7)	4.782(13)
6302	5.862(2)	1.500(8)	2.764(14)

A comparison of the numerical values of the energy of doubly degenerate levels that are linear combinations of the $3s$ and $3d$ states, and also of the $4p$ and $4f$ states with $m=0$ calculated by first-, second-, and third-order perturbation theories according to (44), and the data of exact calculations done by the method of B-splines in Ref. 8 fully confirms the validity of (43): the second- and third-order energies yield the lower and upper limits, with the exact value between them. Here the energy value in the third order is much closer to the exact value than that in the first order (provided the perturbation-theory does not break down).

All diamagnetic corrections up to the third order inclusive calculated in Ref. 7 agree with our data except for the third-order corrections for the doubly degenerate linear combination of the $3s$ and $3d$ levels: both the smaller and the larger of the two values attributed by Vaĭnberg *et al.*[7] to the $3s$ and $3d$ states, respectively, yield absolute values of the susceptibility $\chi^{(3)}$ that are smaller than the values in Table V by approximately 10%.

The laws governing the interdependence of the diamagnetic susceptibilities, noted earlier on the basis of the data of Table V, generally remain valid for states with large values of n (Rydberg states), where there is also correlation between the numerical values of the diamagnetic susceptibilities of the first, second, and third orders. Everywhere except for the boundary between the doublet part[3] (degenerate states of opposite parity whose susceptibilities coincides in all three perturbation-theory orders) and the split part (the numerical values of the even and odd states alternate) of the diamagnetic spectrum an increase in $\chi^{(1)}_\lambda$ with the variation of the sublevel number λ is accompanied by an increase in $\chi^{(2)}_\lambda$ and $\chi^{(2)}_\lambda$. The number of doublet states is approximately one-fourth of the entire set of diamagnetic sublevels (one-fifth of the energy band occupied by the set). Above this boundary the pattern changes to the opposite: $\chi^{(3)}_\lambda$ increases faster than $\chi^{(2)}_\lambda$. At the boundary the monotonic nature of the interdependence between the susceptibilities changes: first there is a sharp decrease, which is followed by an increase in relation to the value of higher-order susceptibilities, with the discontinuity becoming more abrupt as the order of susceptibility increases and manifesting itself more vividly for odd states. For the boundary state the inequality $|\chi^{(N)}_\lambda| < |\chi^{(N)}_{\lambda+1}|$ for $N=2,3$ may be violated. The difference between the maximum ($\chi^{(N)}_K$) and minimum ($\chi^{(N)}_1$) absolute values of the susceptibilities in the diamagnetic set increases with n, so that for a state with $m=0$ the ratio of these values agrees with the approximate formula given above: $\chi^{(N)}_K/\chi^{(N)}_1 \approx 2^{N-1}n$.

As an illustration of the above facts, in Fig. 1 we depict the diagrams representing the interdependence between the absolute values of the first-, second-, and third-order susceptibilities of the diamagnetic sublevels with $n=40$ and $m=0$. Up to $\lambda=5$ the numerical values of the susceptibilities of the even and odd states are almost the same in all

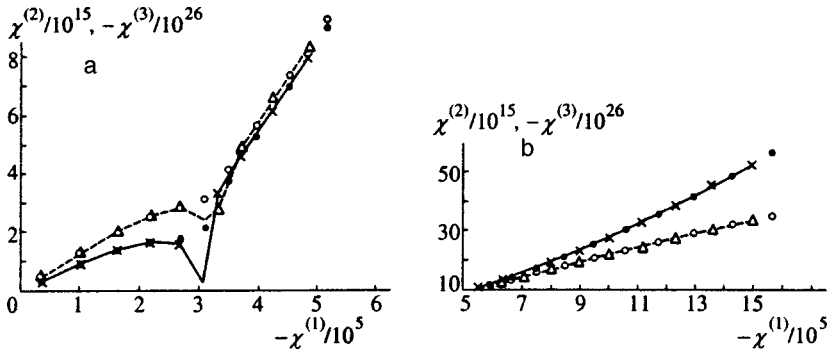

FIG. 1. The interdependence of the first-, second-, and third-order susceptibilities for the first $\lambda=1-10$ (a), and second, $\lambda=11-20$ (b), halves of the diamagnetic states of the level with $n=40$ and $m=0$. The symbols $\triangle$ (connected by dashed curves) and $\bigcirc$ correspond to second-order susceptibility, and the symbols $\times$ (connected by solid curves) and $\bullet$ correspond to third-order susceptibility for even states ($\triangle$ and $\times$) and odd states ($\bullet$).

three orders. An abrupt discontinuity is observed at $\lambda=6$, after which the absolute values of all the susceptibilities of odd states (in Fig. 1 these are connected by solid curves for the third order and by dashed curves for the second order) become smaller than the corresponding susceptibilities of even states. Here, however, the dots representing the dependence of higher-order (second or third) susceptibility on lower-order susceptibility of the even and odd states for $\lambda>7$ lie on the same curve (see Fig. 1a).

The absolute values of the susceptibilities in Fig. 1 show, in particular, that $B\approx 4$T is the limit above which one cannot use perturbation-theory techniques to calculate the diamagnetic energy of states with $n=40$. In such a field the relationship (42) for states of maximum susceptibility ($\lambda=20$ and $p=0,1$) becomes

$$|E_n|:\Delta E^{(1)}_{nm\lambda p}:|\Delta E^{(2)}_{nm\lambda p}|:\Delta E^{(3)}_{nm\lambda p}\approx 1:0.72:0.63:0.62.$$

6. CONCLUSION

The method we have developed for calculating third-order diamagnetic corrections to the energy of hydrogenlike states allowed us to obtain the most complete solution in comparison to the results of earlier calculations (see, e.g., Refs. 6 and 12). In addition to general formulas, which make it possible to find in a fairly simple manner the numerical values of the diamagnetic susceptibilities for arbitrary states, we have extracted new information concerning the structure of the diamagnetic spectrum of highly excited Rydberg atoms. In particular, we have found two features specific to the interdependence of susceptibilities of different orders: a discontinuity, which separates two regions of monotonic increase of susceptibility, at the boundary between the doublet and split diamagnetic states (the discontinuity manifests itself more vividly for higher-order susceptibilities and for states of negative parity), and a variation in the relative rate of monotonic increase of susceptibility in the transition between the two characteristic regions of the diamagnetic spectrum.

The method that we employed to calculate the wave functions and energy in higher-order perturbation theory (see Sec. 2) is fairly general, and the same concept can be used to calculate higher-order perturbation-theory effects in a field that alters the symmetry of degenerate states in such a way that constants of motion of an unperturbed atom cease to exist.

We believe that there is a real possibility of using this method to calculate diamagnetic susceptibilities of higher orders, as it was done by Alliluev and Malkin[19] (see also Ref. 4) in their calculations of the Stark effect, since both the calculation of matrix elements via the Sturm expansion (21) of the Green's function and finding expressions for susceptibilities of the form (14) and (15), together with solving a system of equations of the form (13), can be fully automated by computers.

The work was sponsored by the Russian Fund for Fundamental Research (Grant No. 97-02-16407) and the Russian Ministry of Education (Grant No. 97-0-5.1-63).

*)E-mail: vit@ovd.vsu.ru

1)We use the atomic system of units, in which $e=m=\hbar=1$, with the magnetic field **B** expressed in units of $B_0=2.35\times 10^5$T.

[1] V. S. Lisitsa, Usp. Fiz. Nauk **153**, 379 (1987) [Sov. Phys. Usp. **30**, 927 (1987)].
[2] H. Friedrich, *Theoretical Atomic Physics*, Springer-Verlag, Berlin (1991).
[3] P. A. Braun, Rev. Mod. Phys. **65**, 115 (1993).
[4] H. J. Silverstone, Phys. Rev. A **18**, 1853 (1978).
[5] T. P. Grozdanov and H. S. Taylor, J. Phys. B **19**, 4075 (1986).
[6] M. R. M. Witwit and J. P. Killingbeck, J. Phys. B **26**, 1599 (1973).
[7] V. M. Vaĭnberg, V. A. Gani, and A. E. Kudryavtsev, Zh. Éksp. Teor. Fiz. **113**, 550 (1998) [JETP **86**, 305 (1998)].
[8] J.-H. Wang and C.-S. Hsue, Phys. Rev. A **52**, 4508 (1995).
[9] Yu. P. Kravchenko, M. A. Liberman, and B. Johansson, Phys. Rev. A **54**, 287 (1996).
[10] E. A. Solov'ev, Zh. Éksp. Teor. Fiz. **82**, 1762 (1982) [Sov. Phys. JETP **55**, 1017 (1982)].
[11] T. P. Grozdanov, L. Andric, C. Manescu, and R. McCarroll, Phys. Rev. A **56**, 1865 (1997).
[12] V. D. Ovsiannikov, Phys. Rev. A **57**, 3719 (1998).
[13] N. L. Manakov, V. D. Ovsyannikov, and L. P. Rapoport, Phys. Rep. **141**, 319 (1986).
[14] V. D. Ovsiannikov and S. V. Goossev, Phys. Scr. **57**, 56 (1998).
[15] S. V. Goossev and V. D. Ovsiannikov, J. Phys. B **28**, 5251 (1995).
[16] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977), § 39.
[17] A. Erdélyi, *Higher Transcendental Functions* (Bateman Project) Vol. 1, McGraw-Hill, New York (1953), Chap 5; Vol. 2, Chap. 10.
[18] D. Delande and J. C. Gay, J. Phys. B **17**, L335 (1984).
[19] S. P. Alliluev and I. A. Malkin, Zh. Éksp. Teor. Fiz. **66**, 1283 (1974) [Sov. Phys. JETP **39**, 627 (1974)].

Translated by Eugene Yankovsky

Stimulated superradiance

R. A. Ismailov and A. Ya. Kazakov*)

St. Petersburg State University of Aerospace Instrumentation, 190000 St. Petersburg, Russia
(Submitted 17 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 858–869 (September 1999)

If N atoms simultaneously interact with quasiresonant classical and quasiresonant quantized fields, the modes exchange photons. This processes exhibits cooperative properties, i.e., the number of photons in the quantized mode oscillates, and the amplitude of these oscillations is proportional to N^2. © *1999 American Institute of Physics.* [S1063-7761(99)00809-4]

1. INTRODUCTION

Suppose that a two-level atom simultaneously interacts with the classical and quantized modes that are in quasiresonance with the atomic transition frequency. As shown in Refs. 1–3, the classical and quantized modes exchange photons in such a system. Of course, a similar exchange takes place when N identical atoms simultaneously interact with both modes. The natural question is: Will cooperative properties manifest themselves in such a process? In this paper we show that the answer is yes. More precisely, the dynamics of a system consisting of N atoms, a classical field, and a quantized field exhibits properties similar to those of superradiance.[4] For instance, the number of photons in the quantized mode oscillates, and the amplitude of the oscillations is proportional to N^2. However, in contrast to ordinary superradiance, the oscillation frequency is independent of N.

The physical model of such a system can be realized by a chain of atoms. Here the direction of propagation of the classical field and the axis of the cavity containing the quantized mode are assumed to be perpendicular to the linear chain of atoms. It is also assumed that, on the one hand, the atoms are far enough from each other so that the interaction between them can be ignored but, on the other, are close enough so that they interact with the classical and quantized field in the same way.

To examine this problem we use the Schrödinger-equation picture. Our approach is based on a modified Jaynes–Cummings model. Naturally, the effect of the classical component of the field on the atoms is much more important than that of the quantized component. The evolution of the atoms can be separated into two parts: the ''fast'' evolution associated with the interaction between the atoms and the classical component, and the ''slow'' evolution associated with the interaction between the atoms and the quantized component. In other words, we assume (and this assumption is quite natural) that R, the Rabi parameter of the classical field, is much larger than R_q, the effective Rabi parameter of the quantized mode. This condition makes it possible to separate the fast part of the dynamics of the atoms from the slow part. To this end we use an appropriate averaging procedure (for more details see Refs. 3 and 5) and derive an averaged (over the fast oscillations) Hamiltonian that controls the slow evolution of the system. (From the formal viewpoint, we use an appropriate version of perturbation theory and construct the leading term in the expansion of the solution of the initial-value problem in the small parameter R_q/R.) In physically interesting situations, the averaged Hamiltonian has a simple structure: it is the product of the one-dimensional Fock operator and a matrix operator. This fact makes it possible to reduce the problem of finding the solution for any N to that of solving a set of one-dimensional problems (actually, to solving a similar problem for $N=1$). Therefore, we begin with a description of the results presented in Ref. 3 that are important for us. That paper discusses the case $N=1$. We then use these results to examine the case with an arbitrary number of atoms.

2. THE CASE $N=1$

2.1. General considerations

We begin by briefly discussing the case $N=1$. The starting Hamiltonian is

$$H=\omega a^{\dagger}a+\kappa J_0+\zeta(a^{\dagger}J_-+aJ_+)+\mu[J_-\exp\{i\Omega t\}+J_+\exp\{-i\Omega t\}], \tag{1}$$

where $a^{\dagger}$ and a are the quantized-mode creation and annihilation operators, ω is the frequency of this mode, Ω is the frequency of the classical mode, and the parameter μ we call, at the cost of a certain ambiguity, the Rabi parameter of the classical field. The matrices

$$J_0=\mathrm{diag}\{1,-1\},\quad J_-=J_+^T=\begin{pmatrix}0&0\\1&0\end{pmatrix}$$

describe the two-level atom and its interaction with an external quasiresonant field. The operators and the matrices obey the following relationships:

$$[a,a^{\dagger}]=1,\quad [J_0,J_-]=-2J_-,\quad [J_0,J_+]=2J_+.$$

The parameter ζ in (1) characterizes the interaction between the atom and the quantized mode. The wave function of the system obeys the Schrödinger equation

$$i\frac{\partial\Psi}{\partial t}=H\Psi. \tag{2}$$

We have written the Hamiltonian (1) in the rotating wave approximation. Thus, we assume that the optical fre-

quency is much higher than all other frequency parameters of our problem. Such models (in particular, models of the $\omega=\Omega$ or $\omega=2\kappa=\Omega$ type) were used by Law and Eberly,[6] Alsing *et al.*,[7] and Jyotsna and Agarwal[8] in another physical context, namely, the quantized mode was considered the probe field for the states of the atom.

A remark is in order. We assume that the classical field is monochromatic. As shown in Ref. 3, the use of a polychromatic (i.e., a multifrequency field with equidistant harmonics) does not give rise to new situations with respect to the case of a monochromatic field.

2.2. Averaging over fast oscillations

Our immediate goal is to derive a Hamiltonian, averaged over the fast oscillations, that controls the slow part of the system evolution. We begin with the well-known transformation,

$$\Psi(t)=\exp\left\{-i\omega t\left(a^{\dagger}a+\frac{J_0}{2}\right)\right\}\Phi(t), \tag{3}$$

which isolates the optical frequency, so that

$$i\frac{\partial\Phi}{\partial t}=\left\{\left(\kappa-\frac{\omega}{2}\right)J_0+\zeta(a^{\dagger}J_-+aJ_+)+\mu[J_-\exp\{-i(\omega-\Omega)t\}+J_+\exp\{i(\omega-\Omega)t\}]\right\}\Phi. \tag{4}$$

Now suppose that a 2-by-2 matrix $\Xi(t)$ is the solution of the initial-value problem

$$i\frac{d\Xi}{dt}=\left\{\left(\kappa-\frac{\omega}{2}\right)J_0+\mu[J_-\exp\{-i(\omega-\Omega)t\}+J_+\exp\{i(\omega-\Omega)t\}]\right\}\Xi. \tag{5}$$

$$\Xi(0)=I, \tag{6}$$

where I is the 2-by-2 identity matrix. We can easily write an explicit expression for this matrix:

$$\Xi(t)=\exp\left\{i(\omega-\Omega)t\frac{J_0}{2}\right\}U\exp\{-iRtJ_0\}U^{-1},$$

$$U=\begin{pmatrix}\mu & \Delta-R\\ R-\Delta & \mu\end{pmatrix},\quad R=\sqrt{\mu^2+\Delta^2},\quad \Delta=\kappa-\frac{\Omega}{2}.$$

We seek the solution of Eq. (4) in the form

$$\Phi(t)=\Xi(t)\varphi(t). \tag{7}$$

Using (5), we arrive at equation for $\varphi(t)$:

$$i\frac{\partial\varphi(t)}{\partial t}=\zeta\Xi^{-1}(t)(a^{\dagger}J_-+aJ_+)\Xi(t)\varphi(t). \tag{8}$$

Thus, we can pass from the wave function $\Phi(t)$ to $\varphi(t)$ via the substitution

$$\Phi(t)=\exp\left\{-i\omega t\left(a^{\dagger}a+\frac{J_0}{2}\right)\right\}\Xi(t)\varphi(t),$$

with $\varphi(0)=\Phi(0)$, as (6) implies. Equation (8) is simpler than (2), since its right-hand side is proportional to the small parameter R_q. This makes it possible to select an appropriate asymptotic procedure to build its solution.

What is the physical meaning of the above manipulations? The matrix $\Xi(t)$ is the solution of the problem of the evolution of a two level atom placed in a classical field. Here the function $\varphi(t)$ describes the slow part of the system evolution due to the interaction between the atom and the quantized field. Thus, (7) is an expression for our wave function in the form of the product of the ''fast'' and ''slow'' cofactors. In other words, it implies that we represent our wave function via the basis of atomic states ''dressed by the classical field.''

If we ignore the fast oscillations, which are due to the action of the classical field, we arrive at the following equation:

$$i\frac{\partial\varphi(t)}{\partial t}=\zeta\langle\Xi^{-1}(t)(a^{\dagger}J_-+aJ_+)\Xi(t)\rangle\varphi(t). \tag{9}$$

Here $\langle\cdots\rangle$ indicates that fast oscillations have been ignored. It is here that we invoke the limit $R_q\ll R$ (implicitly; for a more exact description see Refs. 3 and 5).

Simple calculations (omitted here) show that the only harmonics that the operator $\Xi^{-1}(t)(a^{\dagger}J_-+aJ_+)\Xi(t)$ contains are those with the frequencies $\pm(\omega-\Omega)$ and $\pm(\omega-\Omega\pm 2R)$. If these frequencies are of order R, the harmonics are fast and should be discarded in the averaging process, so that only slow harmonics are left. We consider the situation where the averaging yields a nontrivial effect: $|\omega-\Omega|=|2\nu|\ll R$. Leaving only these slow harmonics on the right-hand side of Eq. (9), we arrive at a formula for the leading term in the asymptotic expansion in the small parameter $(\nu,R_q)/R$:

$$H_{\rm av}=\rho(a^{\dagger}\exp\{2i\nu t\}+a\exp\{-2i\nu t\})UJ_0U^{-1},$$

where $\rho=\zeta\mu(R-\Delta)/D$, with $D=\mu^2+(R-\Delta)^2$. The formula implies that in this case, after an expansion in the basis of the matrix UJ_0U^{-1}, the dynamics of the system ''splits'' into two one-dimensional problems.

There are two other cases where the averaging on the right-hand side of Eq. (9) yields a nonzero contribution, with the selection of the slow harmonics being different. This leads to the Jaynes–Cummings standard-model Hamiltonian. In this paper we do not discuss the respective situations.

2.3. A particular case

Let us first consider one of the one-dimensional problems we have just mentioned. We would like to establish the dynamics of the system controlled by the Hamiltonian

$$H_1=\rho(a^{\dagger}\exp\{2i\nu t\}+a\exp\{-2i\nu t\}).$$

We employ the Fock–Bargmann representation,[9] so that $a^{\dagger}\to z$ and $a\to D_z$, with D_z standing for differentiation with respect to z. In this way we reduce this one-dimensional problem to a partial differential equation:

$$i\frac{\partial\varphi(z,t)}{\partial t}=\rho(z\exp\{2i\nu t\}+\exp\{-2i\nu t\}D_z)\varphi(z,t). \tag{10}$$

The solution of Eq. (10) is given by the formula

$$\varphi(z,t)=\exp\left\{\frac{z\rho}{2\nu}(1-\exp\{2i\nu t\})+\frac{\rho^2}{4\nu^2}(\exp\{-2i\nu t\}+2i\nu t-1)\right\}Q\left(z+\frac{\rho}{2\nu}[\exp\{-2i\nu t\}-1]\right). \tag{11}$$

Here the function $Q(z)$ describes the initial distribution of the photons in the quantized mode: $Q(z)=\Phi_0(z)$. For $\nu=0$ Eq. (11) becomes

$$\varphi(z,t)=\exp\left\{-i\rho zt-\frac{\rho^2t^2}{2}\right\}Q(z-i\rho t).$$

According to Ref. 10, for the operator $G=(a^\dagger a)^m$ we have

$$\langle G\rangle=\int dzd\overline{z}\exp\{-z\overline{z}\}\overline{\varphi(z,t)}G\varphi(z,t). \tag{12}$$

Equations (11) and (12) make it possible to calculate the dynamics of any physical variables for any initial data. Let us discuss the situation where the quantized mode in the initial state contains exactly m photons and the state of the atom is an eigenvector of the matrix UJ_0U^{-1}. The eigenvalues of this matrix are ±1. For the case where the eigenvalue is 1 we obtain[1)]

$$\varphi(z,t)=\frac{1}{\sqrt{m!}}\exp\left\{\frac{z\rho}{2\nu}(1-\exp\{2i\nu t\})+\frac{\rho^2}{4\nu^2}(\exp\{-2i\nu t\}+2i\nu t-1)\right\}\left\{z+\frac{\rho}{2\nu}(\exp\{-2i\nu t\}-1)\right\}^m. \tag{13}$$

For the number of photons in the quantized mode we arrive at the following expression:

$$\langle n(t)\rangle=\langle a^\dagger a\rangle=m+\frac{\rho^2\sin^2\nu t}{\nu^2}.$$

At $\nu=0$, i.e., when the frequencies of the classical and quantized fields coincide, we find that $\langle n(t)\rangle=m+\rho^2t^2$.

Our results make it possible to calculate the quantum-statistical characteristics of quantized radiation and to examine the case where the initial state is coherent.[3]

2.4. The general case

Now we discuss the case where the initial state is not an eigenvector of UJ_0U^{-1}. We can easily calculate the eigenvectors of this operator that are orthogonal to each other:

$$e_1=\frac{1}{\sqrt{D}}\begin{pmatrix}\mu\\ R-\Delta\end{pmatrix},\quad e_2=\frac{1}{\sqrt{D}}\begin{pmatrix}\Delta-R\\ \mu\end{pmatrix},$$

where $\varepsilon_k=(-1)^{k+1}$, $k=1,2$, are the corresponding eigenvalues. We seek the solution of the Schrödinger equation with the Hamiltonian $H_{\rm av}$ in the form of a linear combination $\varphi(t)=\eta_1(t)e_1+\eta_2(t)e_2$, where the values of the functions $\eta_k(t)$, $k=1,2$, belong to the Fock space. For these functions we have the following analog of Eq. (10):

$$i\frac{\partial\eta_k(t)}{\partial t}=\varepsilon_k\rho(a^\dagger\exp\{2i\nu t\}+a\exp\{-2i\nu t\})\eta_k(t).$$

The solution of these equations (in the Fock–Bargmann representation) was described earlier, so that

$$\eta_k(z,t)=\exp\left\{\frac{\varepsilon_k\rho z}{2\nu}(1-\exp\{2i\nu t\})+\frac{\rho^2}{4\nu^2}(\exp\{-2i\nu t\}+2i\nu t-1)\right\}Q_k\left(z+\frac{\varepsilon_k\rho}{2\nu}[\exp\{-2i\nu t\}-1]\right),$$

where the function $Q_k(z)$ can be found from the initial data. If

$$\varphi(z,0)=\begin{pmatrix}\varphi_1(z)\\ \varphi_2(z)\end{pmatrix},$$

then

$$Q_1(z)=\frac{\mu\varphi_1(z)+(R-\Delta)\varphi_2(z)}{\sqrt{D}},$$

$$Q_2(z)=\frac{\mu\varphi_2(z)+(\Delta-R)\varphi_1(z)}{\sqrt{D}}.$$

The value of the operator $G=(a^\dagger a)^m$, which acts in the Fock space, can be calculated by the following relationship:

$$\langle G\rangle=\sum_{k=1,2}\int dzd\overline{z}\exp\{-z\overline{z}\}\overline{\eta_k(z,t)}G\,\eta_k(z,t).$$

3. THE CASE $N>1$

3.1. General considerations

Suppose that our system contains $N>1$ identical two-level atoms. For the state space we take $L=F\otimes C^2\otimes C^2\cdots\otimes C^2$, where the Fock space F describes the states of the quantized mode, and N copies of C^2 describe the states of the atoms. Thus, L is the set of linear combinations of vectors of type $f|\nu_1\nu_2\cdots\nu_N\rangle$, where $f\in F$, and the ν_k, $k=1,2,\ldots,N$, are two-dimensional vectors (at the cost of a slight ambiguity, we say that they are the ''components'' of the vector $|\nu_1\nu_2\cdots\nu_N\rangle$). In $C^2\otimes C^2\cdots\otimes C^2$ we use the basis consisting of the vectors $|e_{k_1}e_{k_2}\cdots e_{k_N}\rangle$, where $k_m=1,2$.

We define the operators $J_0^{(m)}$, $J_\pm^{(m)}$, and $\Xi^{(m)}(t)$, $1\leqslant m\leqslant N$, in the following manner: they act on the mth component of the vector $|\nu_1\nu_2\cdots\nu_N\rangle$ as the operators J_0, $J_\pm$, and $\Xi(t)$, respectively, and do not alter the other components. Then the Hamiltonian of our system can be written

$$H=\omega a^\dagger a+\kappa\sum_{m=1}^N J_0^{(m)}+\zeta\sum_{m=1}^N(a^\dagger J_-^{(m)}+aJ_+^{(m)})+\mu\sum_{m=1}^N(J_-^{(m)}\exp\{i\Omega t\}+J_+^{(m)}\exp\{-i\Omega t\}).$$

It describes the simultaneous interaction of N identical two-level atoms with the quasiresonant classical and quasireso-

nant quantized fields and is written in the rotating wave approximation, so that we can assume (as we did before) that the optical frequencies are much higher than all other frequency parameters of our problem.

We are interested in the dynamics of the wave function $\Psi(t)$, the solution of the corresponding Schrödinger equation. We isolate the optical frequency (as we did before) by using an analog of Eq. (3):

$$\Psi(t)=\exp\left\{-i\omega t\left(a^{\dagger}a+\sum_{m=1}^{N}\frac{J_0^{(m)}}{2}\right)\right\}\Phi(t).$$

For the function $\Phi(t)$ we have the equation

$$i\frac{\partial\Phi(t)}{\partial t}=\left\{\left(\kappa-\frac{\omega}{2}\right)\sum_{m=1}^{N}J_0^{(m)}+\zeta\sum_{m=1}^{N}(a^{\dagger}J_-^{(m)}+aJ_+^{(m)})+\mu\sum_{m=1}^{N}[J_-^{(m)}\exp\{i(\Omega-\omega)t\}+J_+^{(m)}\exp\{i(\omega-\Omega)t\}]\right\}\Phi(t).$$

The matrices $\Xi^{(m)}(t)$ (introduced earlier) commute. We introduce the matrix $\Xi_N(t)=\Pi_{m=1}^{N}\Xi^{(m)}(t)$. Obviously, $\Xi_N(0)$ is the identity matrix in $C^2\otimes C^2\cdots\otimes C^2$. The matrix $\Xi_N(t)$ is the solution of the equation

$$i\frac{d\Xi_N(t)}{dt}=\left\{\left(\kappa-\frac{\omega}{2}\right)\sum_{m=1}^{N}J_0^{(m)}+\mu\sum_{m=1}^{N}[J_-^{(m)}\times\exp\{i(\Omega-\omega)t\}+J_+^{(m)}\times\exp\{i(\omega-\Omega)t\}]\right\}\Xi_N(t).$$

Substituting $\Phi(t)=\Xi_N(t)\varphi(t)$, which is the analog of (7) for the case $N>1$, in this equation, we arrive at an equation for $\varphi(t)$:

$$i\frac{\partial\varphi(t)}{\partial t}=\zeta\Xi_N^{-1}(t)\sum_{m=1}^{N}(a^{\dagger}J_-^{(m)}+aJ_+^{(m)})\Xi_N(t)\varphi(t). \tag{14}$$

Note that in our transformations $\varphi(0)=\Psi(0)$. The operator

$$\Xi_N^{-1}(t)\exp\left\{-i\omega t\left(a^{\dagger}a+\sum_{m=1}^{N}\frac{J_0^{(m)}}{2}\right)\right\}\Xi_N(t)$$

is unitary and commutes with $a^{\dagger}a$. Hence for the operator $G=(a^{\dagger}a)^m$ acting in the Fock space F we have $\langle G\rangle=\langle\overline{\Phi(t)},G\Phi(t)\rangle=\langle\overline{\varphi(t)},G\varphi(t)\rangle$.

3.2. The averaging procedure

As in the case with a single atom, the right-hand side of Eq. (14) contains both fast oscillations due to the interaction between the system of atoms and the classical field and slow oscillations due to the interaction with the quantized mode. Discarding the fast harmonics, we arrive at the equation

$$i\frac{\partial\varphi(t)}{\partial t}=\zeta\left\langle\Xi_N^{-1}(t)\sum_{m=1}^{N}(a^{\dagger}J_-^{(m)}+aJ_+^{(m)})\Xi_N(t)\right\rangle\varphi(t).$$

As before, we examine the case $|\omega-\Omega|=|2\nu|\ll R$. This condition imposes a restriction on the difference of the frequencies of the classical and quantized components. Averaging yields a nontrivial contribution, and we find that

$$H_{\rm av}=\rho(a^{\dagger}\exp\{2i\nu t\}+a\exp\{-2i\nu t\})\sum_{m=1}^{N}U_mJ_0^{(m)}U_m^{-1},$$

where the matrix U_m acts (as matrix U) on the mth component and does not alter the other components. Note that the terms on the right-hand side of this expression commute. This Hamiltonian controls the slow evolution of atoms ''dressed by the field'' for the case $N>1$. As in the case $N=1$, the averaged Hamiltonian is the product of the Fock operator $a^{\dagger}\exp\{2i\nu t\}+a\exp\{-2i\nu t\}$ by a matrix. Thus, if we pass to the expansion in the basis of the eigenvectors of this matrix, the averaged Hamiltonian is a set of one-dimensional Hamiltonians. We expand the desired function in this basis: $\varphi(t)=\Sigma_\sigma\eta_\sigma(t)|e_{k_1}e_{k_2}\cdots e_{k_N}\rangle$, where the values of the function $\eta_\sigma(t)$ belong to the Fock space, σ stands for the set of N numbers $k_1,k_2,\ldots,k_N$ each of which is either 1 or 2, and the sum is over all such sets σ (the are 2^N variants in all). For each $\eta_\sigma(t)$ we obtain

$$i\frac{\partial\eta_\sigma(t)}{\partial t}=\rho S_\sigma(a^{\dagger}\exp\{2i\nu t\}+a\exp\{-2i\nu t\})\eta_\sigma(t), \tag{15}$$

where $S_\sigma=\Sigma_{m=1}^{N}(-1)^{k_m+1}$, with the sum incorporating the numbers k_m that comprise the set σ. Here S_σ is the sum over m of the eigenvalues of the operators $U^mJ_0^{(m)}U_m^{-1}$ (the eigenvalues are ±1). What is important here is that of the entire set σ only the factor S_σ is present in S_σ. The solution of the corresponding initial-value problem with $\eta_\sigma(z_0)=Q_\sigma(z)$ has the form $\eta_\sigma(z,t)=\eta_\sigma(z,0)\varphi_\sigma(z,t)$, where

$$\varphi_\sigma(z,t)=\exp\left\{\frac{S_\sigma z\rho\nu}{2}(1-\exp\{2i\nu t\})+\frac{S_\sigma^2\rho^2}{4\nu^2}\times(\exp\{-2i\nu t\}+2i\nu t-1)\right\}Q_\sigma\left(z+\frac{S_\sigma\rho}{2\nu}\times(\exp\{-2i\nu t\}-1)\right).$$

If the operator $G=(a^{\dagger}a)^m$ acts only on the Fock component, we obtain, using the Fock–Bargmann representation,

$$\langle G\rangle=\langle\overline{\varphi(z,t)},G\varphi(z,t)\rangle=\sum_\sigma\int dzd\overline{z}\exp\{-z\overline{z}\}\overline{\eta_\sigma(z,t)}G\eta_\sigma(z,t).$$

These relationships make it possible to describe analytically the solution of any initial-value problem.

3.3. The initial-value problem

Here we discuss the problem corresponding to the following initial data: the quantized mode contains m photons, all atoms are in a single state $v\in C^2$, with $v=e_1\cos\chi$

$+e_2 \sin\chi$. Expanding our wave function in the basis described earlier, we arrive at the following expression for the corresponding coefficients:

$$\eta_\sigma(z,0)=(\cos\chi)^{N-k}(\sin\chi)^k\,\frac{z^m}{\sqrt{m!}},$$

where k is the number of ''twos'' in the set σ. Thus, $\eta_\sigma(z,t)=(\cos\chi)^{N-k}(\sin\chi)^k\varphi_\sigma(z,t)$.

For a given set σ, we denote the number of ''twos'' in the set by $|\sigma|$. Note that if $|\sigma|=k$, then $S_\sigma=N-2k$, and that for a given k there are $\binom{N}{k}$ different sets. Then the expression for $\varphi_\sigma(z,t)$ can be obtained from (11) by replacing ρ with $(N-2k)\rho$.

For the Fock operator $G=(a^\dagger a)^m$ we have

$$\begin{aligned}\langle G\rangle&=\sum_\sigma\int dz\,d\bar{z}\exp\{-z\bar{z}\}\,\overline{\eta_\sigma(z,t)}\,G\,\eta_\sigma(z,t)\\&=\sum_{k=0}^{N}\sum_{|\sigma|=k}\int dz\,d\bar{z}\exp\{-z\bar{z}\}\,\overline{\eta_\sigma(z,t)}\,G\,\eta_\sigma(z,t)\\&=\sum_{k=0}^{N}\binom{N}{k}(\cos\chi)^{2N-2k}(\sin\chi)^{2k}\int dz\,d\bar{z}\\&\quad\times\exp\{-z\bar{z}\}\overline{\varphi_\sigma(z,t)}\,G\,\varphi_\sigma(z,t)\Big|_{|\sigma|=k}.\end{aligned}$$

Here we have allowed for the fact that the coefficients $\eta_\sigma(z,t)$ with the same S_σ have the same dynamics. We can therefore use the above relationships to calculate $\langle G\rangle$.

Let us use these results to find the number of photons in the quantized mode. The corresponding integrals in the above expression were calculated earlier. We have

$$\begin{aligned}\langle n(t)\rangle&=\sum_{k=0}^{N}\binom{N}{k}(\cos\chi)^{2N-2k}(\sin\chi)^{2k}\\&\quad\times\left[m+(N-2k)^2\rho^2\frac{\sin^2\nu t}{\nu^2}\right].\end{aligned}$$

In our calculations we used the simple fact that for any integer s,

$$\begin{aligned}&\sum_{k=0}^{N}\binom{N}{k}(N-2k)^s\cos^{2N-2k}\chi\sin^{2k}\chi\\&\quad=(\cos\chi\sin\chi)^N\,\frac{d^s(\exp y+\exp\{-y\})^N}{dy^s}\Big|_{\exp y=\tan\chi}.\end{aligned}$$

The final formula is

$$\langle n(t)\rangle=m+\frac{\rho\sin^2\nu t}{\nu^2}[N(N-1)\cos^2 2\chi+N]. \qquad (16)$$

Note that N is the number of atoms. Thus, the emission power in the quantized mode, proportional to $\langle n(t)\rangle$, is proportional to N^2. Formula (16) demonstrate a cooperative behavior of the atoms in our system typical of superradiance (see the discussion in Ref. 4). Here, however, in contrast to ordinary superradiance, the temporal dynamics is independent of N. The emission power is a periodic function of time, with the period determined by the difference of the frequencies of the classical and quantized components.

4. CONCLUSION

Our main findings are the following. If a two-level atom simultaneously interacts with a classical field and a quantized mode, the classical and quantized components exchange photons, with the atom acting as a sort of carrier for the photons. If there are N identical two-level atoms simultaneously interacting with the classical and quantized modes and if all the atoms have the same initial state, the photon exchange process exhibits cooperative properties: the emission power is proportional to N^2. Here the temporal dynamics is independent of N and is periodic, with the period determined by the difference of the frequencies of the classical and quantized components.

Let us recall the essence of the superradiance effect.[4] If we have a set of two-level atoms that were initially in the upper state and interact with a quantized field, the interaction is of a collective nature: the number of photons in the quantized mode exhibits a spike whose amplitude is proportional to N^2. If the atoms were to ''discharge'' the photons independently, the number of photons in the quantized mode would be proportional to N. This difference makes it possible to speak of superradiance. For reasons of convenience, the common approach to examining this phenomenon in quantum terms is to study the dynamics of the difference in populations. Then, using the fact that the ''number of excitations'' in such a system is conserved, one can calculate the emission power (see Ref. 4, p. 22). In our case in addition to a classical field acting on the atoms there is a quantized field. Here there is no analog of the law of conservation of the number of excitations. An atom interacting with a classical field can be interpreted as an atom ''dressed'' by the field. Thus, what we are dealing with is an ensemble of atoms ''dressed by the field'' that interact with a quantized field. The results of Ref. 3 imply that there is an exchange of photons between the quantized and classical modes. In the course of this exchange, individual atoms interact via the state of the field in the quantized mode. This interaction has an effect on the transfer of photons, with the process being of a collective nature, and there is a certain analogy between this process and interference phenomena. While in ordinary superradiance the source of photons in the mode is the ensemble of atoms that has been excited in advance, in our case the source of photons in the quantized mode is the classical field. Here the classical field determines the dynamics of the process, which differs dramatically from the dynamics of ordinary superradiance. The very statement of the problem provides the means for interpreting the effect of collective transfer of photons into the quantized mode in an external classical field as stimulated superradiance.

Our results were obtained by a method in which a certain averaging procedure is applied to the initial Hamiltonian, which amounts to using an appropriate version of perturbation theory. The method can be applied under the following conditions: $R_q/R\ll 1$ and $\nu/R\ll 1$. We construct the leading term in the asymptotic expansion of the solution in these

small parameters. The conditions have clear physical meaning and can be easily realized. The first condition means that the amplitude of the quantized field is much smaller than that of the classical field, while the second imposes certain restrictions on the difference of the frequencies of the classical and quantum fields. It is precisely in these conditions that the averaged Hamiltonian, which controls the ''slow'' dynamics of the system, has a simple structure: it is the product of the one-dimensional Fock operator and a purely matrix operator. This fact makes it possible to describe the solution of the problem in explicit analytical terms. Of course, the total Hamiltonian of the problems, which includes correction to the averaged Hamiltonian, does not have such a simple structure, but its terms with a much more complicated structure influence the lower-order terms in the asymptotic expansion of the solutions (for details see Refs. 3 and 5). The structure of the Hamiltonian simplifies considerably if the Hamiltonian is averaged by a procedure based on the results of the well-known work of Poincaré, Bogolyubov, and Mitropol'skiĭ (see Ref. 11).

Thus, we have established the presence of superradiance with rigorous assumptions concerning the initial conditions imposed on the atoms and the fact that the atoms must be identical. Of course, in real situations these assumptions hold only approximately. There is reason to believe, however, that if the necessary corrections are taken into account (within the appropriate version of perturbation theory), the effect will manifest itself. An attractive feature of the model discussed in this paper is that it provides the possibility of calculating in explicit analytical terms all the characteristics of emission in the quantized mode that are of interest from the practical viewpoint.

*)E-mail: akaz@phsc2.stu.neva.ru

1)The results for the case where the eigenvalue is -1 can be obtained by an appropriate change of sign.

[1] A. Ya. Kazakov, Phys. Lett. A **260**, 229 (1995).
[2] A. Ya. Kazakov, Opt. Spektrosk. **81**, 549 (1996) [Opt. Spectrosc. **81**, 498 (1996)].
[3] A. Ya. Kazakov, Quantum Semiclassic. Opt. **10**, 49 (1998).
[4] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics: Superradiation, Bistability, and Phase Transitions*, Nauka, Moscow (1988) [English trans.: *Cooperative Effects in Optics*, IOPP, Bristol (1993)].
[5] A. Ya. Kazakov, Teor. Mat. Fiz. **117**, 92 (1998).
[6] C. K. Law and J. H. Eberly, Phys. Rev. A **43**, 6337 (1991).
[7] P. Alsing, D.-S. Guo, and H. J. Carmichael, Phys. Rev. A **45**, 5135 (1992).
[8] I. V. Jyotsna and G. S. Agarwal, Opt. Commun. **99**, 344 (1993).
[9] A. M. Perelomov, *Generalized Coherent States and their Applications*, Springer-Verlag, New York (1986).
[10] É. I. Aliskenderov, A. S. Shumovskiĭ, and Ho Trung Dung, Fiz. Elem. Chastits At. Yadra **24**, 409 (1993) [Phys. Part. Nuclei **24**, 177 (1993)].
[11] A. H. Nayfeh, *Introduction to Perturbation Techniques*, Wiley, New York (1981).

Translated by Eugene Yankovsky

Classical and quantum versions of the Kapitza–Dirac effect

M. A. Efremov

Moscow Engineering-Physics Institute, 115409 Moscow, Russia

M. V. Fedorov*)

General Physics Institute, Russian Academy of Sciences, 117942 Moscow, Russia
(Submitted 19 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 870–884 (September 1999)

The scattering of electrons on a standing electromagnetic wave (the Kapitza–Dirac effect) is considered within the quantum-mechanical and classical descriptions of the motion of electrons. The mean scattering angle and the distribution function of the electrons after scattering are found. It is shown that assignment of the initial electronic wave function in the form of a plane wave in the quantum-mechanical picture gives rise to the characteristic parameter $\beta = mc^2/\hbar\omega^2\tau$ (where ω and τ are the frequency of the field and the duration of the interaction), which separates the regions of Bragg ($\beta \ll 1$) and nearly classical ($\beta \gg 1$) scattering. When $\beta \gg 1$, the mean scattering angle does not depend on the choice of the description method, but the distribution functions of the electrons after scattering in the quantum and classical approaches differ significantly. When $\beta \ll 1$, both the distribution function and the mean scattering angle differ. Modification of the quantum theory and assignment of the initial electronic wave function in the form of a more or less localized wave packet are apparently needed to eliminate the differences discovered. The results obtained can be used to determine the dimensions of the wave packet which characterizes the state of electrons in a beam from experiments on scattering from a standing light wave. © *1999 American Institute of Physics.* [S1063-7761(99)00909-9]

1. INTRODUCTION

The scattering of electrons on a standing electromagnetic wave was first examined by Kapitza and Dirac in 1933.[1] In the ensuing years, this effect (the Kapitza–Dirac effect) was widely investigated by both theoretical[2–4] and experimental[5–12] methods. The most convincing experimental data were obtained in Ref. 10, although they refer to the range of fairly strong fields ($\sim 10^{13}\,\mathrm{W\cdot cm^{-2}}$) and the effect observed can be interpreted as a generalization of the Kapitza–Dirac effect to the multiphoton case.[2,3]

From the quantum standpoint, the Kapitza–Dirac effect is induced Compton scattering. Since a standing wave is a superposition of two counterpropagating traveling waves of identical frequency (ω), the process of induced Compton scattering in the case under consideration consists of the absorption of a photon from one of the traveling waves and the induced emission of a photon corresponding to the other wave. The energy of the electron clearly does not vary, but its momentum varies by $\pm 2\hbar\mathbf{k}$, where $\mathbf{k}$ is the wave vector of one of the traveling waves. For assigned ω and $\mathbf{k}$ it follows from the condition of equality between the energies of the electron before and after scattering that such a process is possible only for certain directions of the initial momentum of the electron, rather than for any direction. According to Ref. 1, the stimulated Compton scattering of electrons on a standing wave can be interpreted as the diffraction of the de Broglie wave of an electron on the periodic structure with a period $\lambda/2$ (where $\lambda = 2\pi c/\omega$ is the wavelength of the radiation field) formed by the planes of equal phase (antinodes) in the standing wave. The condition just noted, which follows from the energy and momentum conservation laws and determines the direction of motion of the incident electron for which induced Compton scattering on the standing wave is possible, is interpreted as the Wulff–Bragg condition for diffraction of the de Broglie wave of an electron on a periodic lattice.[1] If $\pi/2 - \alpha$ is the angle between the direction of the momentum of the incident electron $\mathbf{p}_0$ and the wave vector $\mathbf{k}$ (α is the glancing angle of the incident electron), the Wulff–Bragg condition has the form $\alpha = \pm\alpha_{\mathrm{Br}}$, where α_{Br} is the Bragg angle:

$$\alpha_{\mathrm{Br}} = \arcsin\left(\frac{\lambda_{\mathrm{DB}}}{\lambda}\right) = \arcsin\left(\frac{\hbar k}{p_0}\right) \approx \frac{\hbar k}{p_0}. \tag{1}$$

Under typical conditions ($\omega = 3\times 10^{15}\,\mathrm{s^{-1}}$ and $v_0 = p_0/m = 10^8\,\mathrm{cm\cdot s^{-1}}$) $\alpha_{\mathrm{Br}} \approx 10^{-3} \ll 1$. Below we shall assume that not only the Bragg angle α_{Br}, but also the initial and final glancing angles α and α', as well as the electron scattering angle $\theta = \alpha' - \alpha$, are small (see Fig. 1).

As far as we know, a quantum-mechanical description of the electron was used in all the theoretical studies of the Kapitza–Dirac effect which have been performed, and its initial wave function was assigned in the form of a plane wave. Nevertheless, another approach, viz., the scattering of a classical electron by the field of a classical standing light wave, is justified and interesting. Such a formulation of the problem is examined in the fourth section of this paper. The

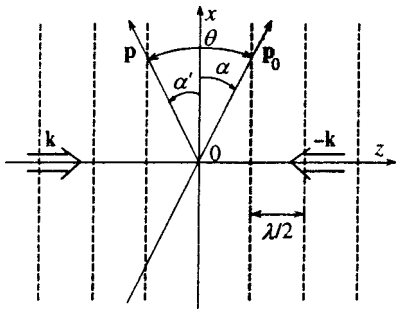

FIG. 1. Scheme for the scattering of electrons on a standing light wave in the Kapitza–Dirac effect.

fundamental approximations and a general statement of the problem are formulated in Sec. 2. The third section provides a brief account of the known and some previously unreported results of the quantum treatment. A comparison of the results of the classical and quantum approaches reveals some significant differences. A discussion of the reasons for these differences and the prospects for future studies are given in the fifth section of this article. We confine ourselves in this paper to the weak-field approximation for the standing wave, and the electron velocity is assumed to be small compared with the velocity of light.

2. GENERAL FORMULATION OF THE PROBLEM

To fix ideas, let the vector $\mathbf{k}$ be directed along the z axis. Also, let the field of the standing wave be linearly polarized and the electric field strength vector $\mathbf{E}$ be directed along the x axis, and let the initial momentum $\mathbf{p}_0$ of the electron lie in the xz plane and form an angle α with the direction of the x axis (Fig. 1). For simplicity, we take a planar geometry, i.e., the radiation field does not depend on the coordinate y and the projection of the momentum of the electron onto the direction of the y axis is equal to zero both before and after scattering.

In both the quantum and classical descriptions of the motion of an electron, the starting point for formulating the problem is its Hamiltonian in the external radiation field. Assuming that in either case the external electromagnetic radiation can be described classically, for a configuration such as a standing wave we assign the electric field strength in the form

$$\mathbf{E}=\mathbf{E}_0(t)[\cos(\omega t-kz)+\cos(\omega t+kz)], \tag{2}$$

where $E_0(t)$ is the amplitude envelope of the radiation pulse and $k=\omega/c$. As is generally known, the Hamiltonian of a nonrelativistic electron in the field (2) has the form

$$H=\frac{(\mathbf{p}-e\mathbf{A}(t)/c)^2}{2m}, \tag{3}$$

where $\mathbf{A}(t)$ is the vector potential corresponding to the electric field strength (2):

$$\mathbf{A}=-\frac{c\mathbf{E}_0(t)}{\omega}[\sin(\omega t-kz)+\sin(\omega t+kz)]. \tag{4}$$

The expression (3) for the Hamiltonian of the electron is equally applicable to both the quantum and classical descriptions with the one difference that while in the classical approach the momentum of the electron $\mathbf{p}=\mathbf{p}(t)$ is an ordinary function, in quantum theory it is the operator $\mathbf{p}=-i\hbar\nabla$.

We use the averaged Hamiltonian approximation.[2–4] In this approximation it is assumed that an electron in a field undergoes both fast and slow motions (on the scale of the field period $2\pi/\omega$), but the amplitudes of the fast changes in the coordinates, momentum, and wave function are small compared with the large-scale slow changes. This assumption allows us to obtain equations for the slow components of the quantities under consideration using simple averaging of the Hamiltonian (3) with respect to time, which gives

$$\overline{H}=\frac{\mathbf{p}^2+e^2\overline{\mathbf{A}^2(t)}/c^2}{2m}\cong H_0+H_{\rm int}. \tag{5}$$

Here $H_0=\mathbf{p}^2/2m$ is the Hamiltonian of the free electron, and $H_{\rm int}$ is the Hamiltonian of its interaction with the field, which is conveniently written in the form

$$H_{\rm int}=2U(t)\cos(2kz), \tag{6}$$

where $U(t)$ is ponderomotive potential of the electron, which depends slowly on time:

$$U(t)=\frac{e^2E_0^2(t)}{4m\omega^2}. \tag{7}$$

It is noteworthy that the problem of electron scattering can be treated in both nonstationary and stationary formulations. If each of the two counterpropagating traveling waves has the form of short pulses and if the pulse duration is short compared with the time needed for an electron to cross the focal region, then, as was pointed out above, $E_0(t)$ is the amplitude envelope of the field pulses. In this case the scattering problem has an explicitly nonstationary character. On the other hand, if the pulse duration is much greater than the time needed for an electron to cross the focal region, the field can be considered stationary, but inhomogeneous: $E_0=E_0(x)$. In this case, the interaction with the field is switched on and switched off as an electron traverses the focal region. It can be seen that in both the quantum and classical descriptions the stationary scattering problem can be reduced to the nonstationary problem under certain, fairly natural assumptions by introducing the ''effective time'' $t=x/v_{0x}$ and changing the notation from $E_0(x)=E_0(tv_{0x})$ to $E_0(t)$, where v_{0x} is the projection of the initial velocity v_0 of the electron onto the x axis. For this reason, we confine ourselves to a treatment of the nonstationary scattering problem, bearing in mind the possibility of using the results in the stationary case.

Below, without explicitly specifying the form of the field envelope $E_0(t)$, we assume that it is defined by a certain even function of time $f(t)$, which is such that

$$f(t)=f(-t),\quad f_{\max}=f(0)=1,\quad E_0(t)=E_0f(t),$$

and, in accordance with Eq. (7),

$$U(t)=U_0f^2(t),$$

where E_0 and $U_0=e^2E_0^2/4m\omega^2$ are the maximum values of the electric field strength amplitude $E_0(t)$ and the ponderomotive potential $U(t)$. The ultimate purpose of solving both the classical and quantum problems will be to calculate the mean electron scattering angle and the angular distribution function of electrons after scattering.

3. MODEL OF PLANE-WAVE SCATTERING IN THE QUANTUM THEORY OF THE KAPITZA–DIRAC EFFECT

3.1. Mean deflection angle of electrons in the field of a standing wave

Using the Hamiltonian (5), we seek a solution of the Schrödinger equation in the form of an expansion in plane waves, which are conveniently normalized to unity in a periodicity cube (its volume is V, and in the final results $V\to\infty$)

$$\Psi(\mathbf{r},t)=\frac{1}{\sqrt{V}}\sum_p C(\mathbf{p},t)\exp\left[\frac{i}{\hbar}\left(\mathbf{p}\cdot\mathbf{r}-\frac{p^2}{2m}t\right)\right]. \tag{8}$$

The expansion coefficients of the wave function $C(\mathbf{p},t)$ satisfy an equation which follows directly from the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}C(\mathbf{p},t)=\sum_{p'} H_{\rm int}(\mathbf{p}'-\mathbf{p}) \times\exp\left[\frac{i}{\hbar}\left(\frac{p^2-p'^2}{2m}\right)t\right]C(\mathbf{p}',t), \tag{9}$$

where $H_{\rm int}(\mathbf{q})$ is the Fourier transform of the interaction energy (6), or, more specifically,

$$H_{\rm int}(\mathbf{q})=\frac{1}{V}\int d\mathbf{r}\,H_{\rm int}(\mathbf{r})\exp\left(\frac{i\mathbf{q}\cdot\mathbf{r}}{\hbar}\right). \tag{10}$$

In the model of plane-wave scattering it is assumed that before an electron begins to interact with the field, it is in a state with a definite momentum $\mathbf{p}_0$, and the initial condition for Eq. (9) has the form

$$C(\mathbf{p},t\to-\infty)=\delta_{p,p_0}. \tag{11}$$

The expansion coefficients in Eq. (8) $C(\mathbf{p},t)$ are the probability amplitudes, and $|C(\mathbf{p},t)|^2$ is the probability of finding an electron in a state with the momentum $\mathbf{p}$ at the time t. The mean change in the projection of the electron momentum onto the z axis as a result of scattering is defined as

$$\langle\Delta p_z\rangle=\sum_p (p_z-p_{0z})|C(\mathbf{p},t\to+\infty)|^2. \tag{12}$$

In the approximation of a small change in the electron momentum $\langle\Delta p_z\rangle\ll p_0\langle\Delta p_z\rangle$ also determines the mean electron scattering angle

$$\overline{\theta}\equiv\langle\alpha'-\alpha\rangle=\left\langle\arcsin\left[\frac{\Delta p_z}{|\mathbf{p}_0+\Delta\mathbf{p}|}\cos(\alpha)\right]\right\rangle \approx\frac{\langle\Delta p_z\rangle}{p_0}=\sum_p\frac{p_z-p_{0z}}{p_0}|C(\mathbf{p},t\to+\infty)|^2. \tag{13}$$

We find the first-order correction $C^{(1)}(\mathbf{p},t=\infty)$ to the unperturbed probability amplitude $C^{(0)}(\mathbf{p},t)\equiv C(\mathbf{p},t=-\infty)=\delta_{p,p_0}$ directly from Eq. (9) in the weak-field approximation within first-order perturbation theory with respect to the interaction of an electron with the field with consideration of the initial condition (11) and the explicit form of the interaction energy $H_{\rm int}$ (6):

$$C^{(1)}(\mathbf{p},\infty)=-\frac{i}{\hbar}U_0\sum_{\pm}\delta_{p,p_0\pm2\hbar k}(f^2)_{\pm}, \tag{14}$$

where $(f^2)_{\pm}$ is the Fourier transform of the square of the dimensionless field envelope $f(t)$

$$(f^2)_{\nu}=\int_{-\infty}^{\infty}f^2(t)\exp(i\nu t)dt, \tag{15}$$

calculated for values of the Fourier-transformation ''frequency'' ν which equal

$$\nu_{\pm}=\frac{(\mathbf{p}_0\pm2\hbar\mathbf{k})^2-p_0^2}{2m\hbar}\approx\pm2\omega\frac{v_0}{c}(\alpha\pm\alpha_{\rm Br}). \tag{16}$$

It is easy to see that the ''frequencies'' $\nu_{\pm}$ specify the proximity of the glancing angle α of the incident electrons to the Bragg angles $\mp\alpha_{\rm Br}$. Taking into account that at small values of ν, $(f^2)_{\nu}\sim\tau$ (where τ is the duration of the interaction) and comparing $C^{(1)}$ from Eq. (14) with $C^{(0)}$ from Eq. (11), we find that the perturbation-theory parameter is $U_0\tau/\hbar$. The condition for applicability of the weak-field approximation in the scattering model under consideration, $U_0\tau/\hbar<1$, is satisfied, for example, for $\omega=3\times10^{15}\,\mathrm{s}^{-1}$ and $\tau=1$ ps and 1 ns when the radiation intensity $I\leqslant10^{10}\,\mathrm{W\cdot cm}^{-2}$ and $10^7\,\mathrm{W\cdot cm}^{-2}$, respectively.

Using Eqs. (13) and (14), we can easily find the mean electron scattering angle in the first order of quantum perturbation theory. Because there are Kronecker deltas in Eqs. (11) and (14), the glancing angle of the scattered electrons in a beam, $\alpha'=\arcsin(p_z/p)$, can take the values $\alpha'=\alpha$ and $\alpha'\approx\alpha\pm2\alpha_{\rm Br}$. In diffraction theory the equality $\alpha'-\alpha=\pm2\alpha_{\rm Br}$ is known as the Laue condition.[13] The mean scattering angle is

$$\overline{\theta}=\left(\frac{U_0}{\hbar}\right)^2 2\alpha_{\rm Br}\{|(f^2)_+|^2-|(f^2)_-|^2\}. \tag{17}$$

In the special case of a Gaussian envelope, $f^2(t)=\exp(-(t/\tau)^2)$, Eq. (17) takes the form

$$\overline{\theta}=\left(\frac{U_0}{\hbar}\tau\right)^2 2\pi\alpha_{\rm Br}\left\{\exp\left(-\frac{\nu_+^2\tau^2}{2}\right)-\exp\left(-\frac{\nu_-^2\tau^2}{2}\right)\right\}. \tag{18}$$

The mean scattering angle defined by Eq. (17) or (18) depends not only on the parameters of the field pulse (intensity, pulse duration, and envelope shape), but also on the direction of the momentum of the electron in its initial state $\mathbf{p}_0$ or, stated differently, on the glancing angle of the incident electron α: $\overline{\theta}=\overline{\theta}(\alpha)$. Plots of this dependence calculated for

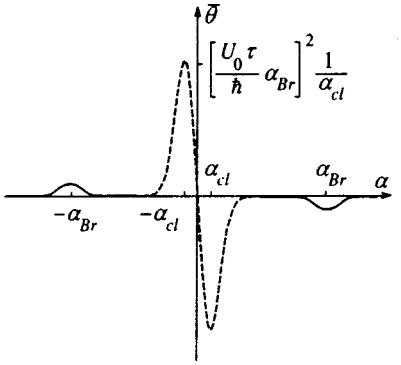

FIG. 2. Plots of $\overline{\theta}(\alpha)$ determined from the quantum (solid curve) and classical (dashed curve) theories for $\beta=0.2$.

the case of a Gaussian pulse are depicted by the solid curves in Figs. 2 and 3. The solid curve in Fig. 2 corresponds to the standard models of the Kapitza–Dirac effect, according to which the mean scattering angle is nonzero only if the initial glancing angle α is close to the Bragg angle α_{Br}. In the vicinity of these values of α, the solid curve in Fig. 2 has the form of narrow peaks with a height $\pm(U_0\tau/\hbar)^2 2\pi\alpha_{\mathrm{Br}}$ and a width

$$\Delta\alpha=\frac{1}{\omega\tau v_0/c}=\alpha_{\mathrm{Br}}\beta\ll\alpha_{\mathrm{Br}},$$

where β is one of the fundamental parameters of the quantum theory of the Kapitza–Dirac effect

$$\beta=\frac{mc^2/\hbar\omega}{\omega\tau}. \tag{19}$$

Both the numerator and the denominator of the fraction on the right-hand side of Eq. (19) are large, i.e., $mc^2/\hbar\omega\gg1$ and $\omega\tau\gg1$, and their ratio (β) can be either large or small. If $\omega=3\times10^{15}\,\mathrm{s}^{-1}$, the value $\beta=1$ corresponds to a duration of the interaction $\tau\approx10^{-10}\,\mathrm{s}$. The curves in Figs. 2 and 3 were obtained for $\beta\ll1$ and $\beta\gg1$, which correspond to nanosecond and picosecond values of τ, respectively, when $\omega=3\times10^{15}\,\mathrm{s}^{-1}$. When $\beta\ll1$, because of the narrow width of the peaks on the $\overline{\theta}(\alpha)$ curve, the terms in the curly brackets on the right-hand sides of Eqs. (17) and (18) do not interfere with one another. They can be nonzero only individually: the first terms at $\alpha\approx-\alpha_{\mathrm{Br}}$ and the second terms at $\alpha\approx\alpha_{\mathrm{Br}}$. As the duration of the interaction decreases (i.e., as β increases) the peaks on the $\overline{\theta}(\alpha)$ curve broaden, and in the case of $\beta\gg1$ the width of each of the peaks becomes considerably greater than the distance between them. Under these conditions, unlike the case of $\beta\ll1$, the first and second terms in Eqs. (17) and (18) can interfere and compensate one another, significantly altering the structure of the $\overline{\theta}(\alpha)$ curve (Fig. 3) and rendering it different from the curve corresponding to the usual models of the Kapitza–Dirac effect. As a whole, at large values of β the peaks on the $\overline{\theta}(\alpha)$ curve become broad and shift into the region $|\alpha|\gg\alpha_{\mathrm{Br}}$. Using this inequality, we can transform Eqs. (17) and (18) and expand the Fourier transforms of the square of the field envelope (15) in powers of α_{Br} in the definition of $\nu_\pm$ (16). Equation (17) thus transformed takes the form

$$\overline{\theta}\approx\left(\frac{U_0}{\hbar}\alpha_{\mathrm{Br}}\right)^2 8\omega\frac{v_0}{c}\left.\frac{d|(f^2)_\nu|^2}{d\nu}\right|_{\nu=2\alpha\omega v_0/c}, \tag{20}$$

where, as before, $(f^2)_\nu$ is defined by Eq. (15). In the special case where $f(t)$ is a Gaussian function, the corresponding result can be obtained either from Eq. (20) or directly from Eq. (18) using the expansion in α_{Br}:

$$\overline{\theta}\approx-\left(\frac{U_0\tau^2\omega}{\hbar}\frac{v_0}{c}\alpha_{\mathrm{Br}}\right)^2 16\pi\alpha\exp\left[-2\left(\omega\tau\frac{v_0}{c}\alpha\right)^2\right]. \tag{21}$$

Unlike (17) and (18), the right-hand sides of Eqs. (20) and (21) no longer depend on Planck's constant $\hbar$ and can be regarded as the classical limit of the quantum theory of the Kapitza–Dirac effect. The result (21) corresponds to the dotted curve in Fig. 3 [we recall that the solid curve corresponds to the exact formula (18)]. The principal parameters of the $\overline{\theta}(\alpha)$ curve specified by (21) are the height of the peaks $\pm\overline{\theta}_{\max}$, their width $\Delta\alpha$, and their positions $\mp\alpha_{\mathrm{cl}}$:

$$\overline{\theta}_{\max}=\left(\frac{U_0}{mv_0c}\omega\tau\right)^2 8\pi\omega\tau\frac{v_0}{c}\exp\left(-\frac{1}{2}\right),$$

$$\Delta\alpha\sim\alpha_{\mathrm{cl}}=\frac{1}{2\omega\tau v_0/c}=\frac{1}{2}\beta\alpha_{\mathrm{Br}}\gg\alpha_{\mathrm{Br}}. \tag{22}$$

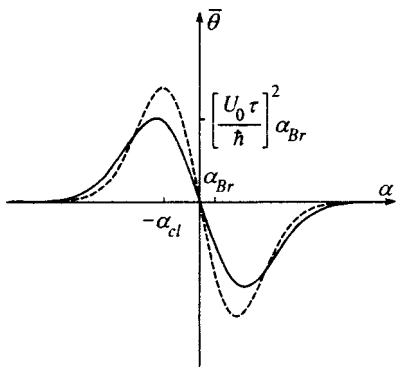

FIG. 3. Plots of $\overline{\theta}(\alpha)$ determined from the quantum (solid curve) and classical (dashed curve) theories for $\beta=2$.

3.2. Angular distribution function of scattered electrons

The mean electron scattering angle is not the only characteristic of the scattering process which can be calculated either by theoretical means or from experimental data. Apparently, the distribution function $F(\theta)$ of the scattered electrons with respect to the angle θ or the number of electrons $F(\theta)d\theta$ detected by the sensor in an assigned direction θ in an interval of angles of width $d\theta$ is always measured directly in an experiment. Within the model of plane-wave scattering considered above, the number of electrons having a momentum in the interval $[\mathbf{p},\mathbf{p}+d\mathbf{p}]$ is related directly to the expansion coefficients $C(\mathbf{r},t)$ of the electronic wave function $\Psi(\mathbf{r},t)$ in plane waves (8):

$$\frac{dw}{d\mathbf{p}}=\frac{V}{(2\pi\hbar)^3}|C(\mathbf{p},t\to+\infty)|^2. \tag{23}$$

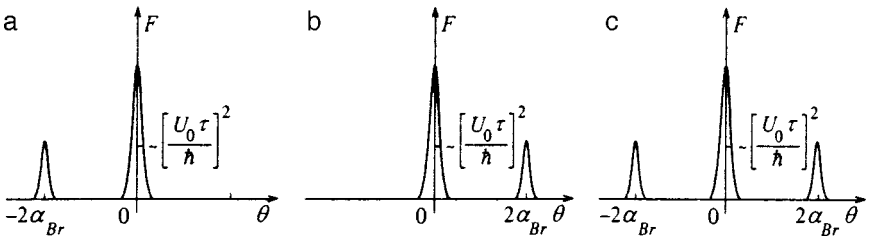

FIG. 4. Distribution function of scattered electrons found from the quantum theory for $\beta \ll 1$, $\alpha = \alpha_{\mathrm{Br}}$ (a); $\beta \ll 1$, $\alpha = -\alpha_{\mathrm{Br}}$ (b); and $\beta \gg 1$ (c).

Because $\Delta p_z \approx p_0 \theta$ and $dp_z \approx p_0 d\theta$ for small momentum changes ($|\Delta p_z| \ll p_0$), the probability density (23) can be related directly to the angular distribution function of the scattered electrons

$$F(\theta) = \frac{dw}{d\theta} = \int d\mathbf{p}_\perp \frac{dw}{d\mathbf{p}_\perp d(p_z/p_0)} = \frac{p_0 V}{(2\pi\hbar)^3} \int d\mathbf{p}_\perp |C(\mathbf{p}, t \to +\infty)|^2 |_{p_z = p_0[\alpha+\theta]}, \tag{24}$$

where $d\mathbf{p}_\perp = dp_x dp_y$. When $C(\mathbf{p},t)$ is calculated from perturbation theory and the expressions (11), (14), etc. are used, the Kronecker deltas of the type $\delta_{p,p'}$ are expressed in terms of Dirac δ functions using relations of the type

$$\delta_{p,p'} = \frac{(2\pi\hbar)^3}{V} \delta(\mathbf{p} - \mathbf{p}'),$$

and, as a result, we find

$$F(\theta) = \left\{ 1 - \left(\frac{U_0}{\hbar}\right)^2 [|(f^2)_+|^2 + |(f^2)_-|^2] \right\} \delta(\theta) + \left(\frac{U_0}{\hbar}\right)^2 [|(f^2)_+|^2 \delta(\theta - 2\alpha_{\mathrm{Br}}) + |(f^2)_-|^2 \delta \times (\theta + 2\alpha_{\mathrm{Br}})], \tag{25}$$

where, as before, the Fourier transforms of the square of the field envelope $(f^2)_\pm$ are given by Eqs. (15) and (16). We note that the appearance of singularities in the distribution function is associated with the unboundedness of the integration region with respect to z. Actually the size of this region can be restricted, for example, by the distance between the mirrors or the length of the focus L. As a result, the δ functions in Eq. (25) and below are replaced by functions of a finite, but small width $\delta\theta \sim \hbar / L p_0$.

The distribution function (25) satisfies the following obvious requirements: it is normalized to unity, i.e.,

$$\int F(\theta) d\theta = 1,$$

and the mean scattering angle calculated from (25),

$$\overline{\theta} = \int \theta F(\theta) d\theta,$$

coincides with the previously derived expression (17). However, of course, the distribution function contains considerably more information on the properties of the scattering process. In particular, using $F(\theta)$, we can easily calculate any moments of the scattering angle, for example, the mean square of the scattering angle

$$\overline{\theta^2} = \int \theta^2 F(\theta) d\theta = \left(\frac{U_0}{\hbar}\right)^2 4\alpha_{\mathrm{Br}}^2 [|(f^2)_+|^2 + |(f^2)_-|^2] \tag{26}$$

etc.

The distribution function (25) is shown in Fig. 4 for three cases: $\beta \ll 1$, $\alpha = \alpha_{\mathrm{Br}}$ (a); $\beta \ll 1$, $\alpha = -\alpha_{\mathrm{Br}}$ (b); and $\beta \gg 1$ (c). In the first two cases there is a single additional peak (in comparison to the principal peak corresponding to the absence of scattering), which appears only when the Bragg condition $\alpha = \alpha_{\mathrm{Br}}$ or $\alpha = -\alpha_{\mathrm{Br}}$ is satisfied. In the third case ($\beta \gg 1$) both additional peaks are represented in the distribution function $F(\theta)$ over a broad range of variation of α (out to $|\alpha| \sim \alpha_{\mathrm{cl}} \gg \alpha_{\mathrm{Br}}$), and their heights are commensurate in magnitude (they are exactly equal in height only in the case of $\alpha = 0$, where we consequently have $\overline{\theta} = 0$).

4. SCATTERING OF A CLASSICAL ELECTRON ON A STANDING WAVE

4.1. Mean scattering angle

Within the classical approach, consideration of the time-averaged Hamiltonian (5) as a classical Hamilton function with the potential energy H_{int} (6) yields a one-dimensional Newtonian equation describing the motion of an electron in the field of a standing wave in the direction $z \| \mathbf{k}$:

$$m \frac{d^2 z(t)}{dt^2} = 4kU(t) \sin(2kz(t)), \tag{27}$$

where $U(t)$ is the ponderomotive potential (7). We assign the initial conditions for Eq. (27) at a certain distant moment in time t_0 before the beginning of the interaction with the field:

$$z(t_0) = z_0, \quad \dot{z}(t_0) = v_{0z}, \tag{28}$$

where z_0 is the initial coordinate and v_{0z} is the initial velocity of the electron parallel to the z axis. At the same time, the velocity of the electron in the perpendicular direction (parallel to the x axis) is identically equal to a constant, and $x(t) \equiv x_0 + v_{0x}(t - t_0)$. As can easily be proved, Eq. (27) is equivalent to an integral equation of the form

$$z(t) = z_0 + v_{0z}(t - t_0) + \frac{4k}{m} \int_{t_0}^{t} (t - t') U(t') \sin(2kz(t')) dt'. \tag{29}$$

We seek a solution of Eq. (29) in the weak-field approximation by iterating with respect to the interaction, i.e., in the form of a series in powers of $U(t)$:

$$z^{(0)}(t)=z_0+v_{0z}(t-t_0),$$

$$z^{(1)}(t)=\frac{4k}{m}\int_{t_0}^{t}(t-t')U(t')\sin(2kz^{(0)}(t'))dt', \tag{30}$$

$$z^{(2)}(t)=\frac{8k^2}{m}\int_{t_0}^{t}(t-t')U(t')z^{(1)}(t')\cos(2kz^{(0)}(t'))dt',$$

etc.

The scattering angle θ of a classical particle is determined by the direction of its velocity after scattering. In the approximation of small scattering angles we can write

$$\theta=\frac{\dot z(t\to\infty)-v_{0z}}{v_0}\cos(\alpha), \tag{31}$$

where $v_0=(v_{0x}^2+v_{0z}^2)^{1/2}$ is the total initial velocity of the electron and α, as before, is the glancing angle before scattering (see Fig. 1). Using the perturbation formulas (30) for the coordinate z and the definition (31), we can easily find the corresponding explicit expressions for the scattering angle. In first-order perturbation theory the result of the calculations has the form

$$\theta^{(1)}(t)=\theta_m(t)\sin(2kz_0-\varphi), \tag{32}$$

where

$$\theta_m(t)=\frac{4\omega U_0}{mv_0c}\left|\int_{t_0}^{t}dt'f^2(t')\exp\left(2i\alpha\omega\frac{v_0}{c}\right)\right|,$$

$$\varphi(t)=2kv_{0z}t_0-\arcsin\left[\int_{t_0}^{t}dt'f^2(t')\sin\left(2\alpha\omega\frac{v_0}{c}t'\right)\Big/\int_{t_0}^{t}dt'f^2(t')\cos\left(2\alpha\omega\frac{v_0}{c}t'\right)\right]. \tag{33}$$

In the case of a smooth field envelope assigned by an even function $f(t)$, Eqs. (33) become significantly simpler when $t\to\infty$ and $t_0\to-\infty$:

$$\theta_m=\frac{4\omega U_0}{mv_0c}|(f^2)_\nu|_{\nu=2\alpha\omega v_0/c},\quad \varphi=2kv_{0z}t_0, \tag{34}$$

where, as before, $(f^2)_\nu$ is the Fourier transform of the square of the envelope $f(t)$ (15).

Under the assigned initial conditions (28), the scattering angle θ (31) is uniquely defined in the classical treatment. A spread of scattering angles appears when we proceed from the treatment of the motion of a single particle to the scattering of a beam of particles, whose values of the initial parameters z_0 and v_{0z} are distributed in some manner. Let $v_{0z}=\text{const}$ in the incident beam, and let the values of the initial coordinate z_0 be distributed in the periodicity interval $\lambda/2$ of the ponderomotive potential (7). By definition, the mean scattering angle of the particles in a beam equals

$$\overline{\theta}_{\rm cl}=\langle\theta(z_0,t\to\infty)\rangle_{z_0}=\left\langle\frac{\dot z(t\to\infty)-v_{0z}}{v_0}\right\rangle_{z_0}$$

$$=\int_0^{\lambda/2}\frac{dz_0}{\lambda/2}\,\frac{\dot z(t\to\infty)-v_{0z}}{v_0}. \tag{35}$$

It is easy to see that in first-order perturbation theory with respect to U_0 the substitution of $\theta^{(1)}$ (32) into Eq. (35) followed by averaging over z_0 gives zero. A nonzero mean scattering angle appears only in the second order with respect to U_0. In view of the cumbersome form of the general formulas, we present only the expression for the second-order scattering angle averaged with respect to z_0

$$\overline{\theta_{\rm cl}^{(2)}}=\frac{8k^3U_0^2}{m^2v_0}\frac{d}{d\nu}|(f^2)_\nu|^2\Big|_{\nu=2\alpha\omega v_0/c}, \tag{36}$$

which is equivalent to the result (20) obtained in the preceding section in the classical limit of the quantum theory of the scattering of a plane wave. It should be noted that in the quantum theory of the scattering of a plane wave, the classical limit (20) is obtained only in the approximation $\beta\gg1$, where β is the parameter defined by Eq. (19). In the classical theory of scattering there is no such constraint in the general case. It can be stated that the classical theory of scattering extends the result (20), which was previously determined as the classical limit of the quantum theory, to the entire range of large and small values of β. Moreover, while the classical treatment and the quantum theory of the scattering of a plane wave give the identical results (20) and (36) in the region $\beta\gg1$, in the opposite case, where $\beta\ll1$, the results of the quantum [see (17) and (18)] and the classical [see (36)] theories differ significantly, as is clearly seen from Fig. 2, where the solid and dashed curves are plots of the forms of $\overline{\theta}(\alpha)$ obtained from the quantum and classical treatments. It seems to us that this difference between the results of the quantum and classical analyses can be eliminated if the scattering of wave packets, rather than plane waves, would be considered in a quantum theory.[14–17] In addition, if the width of the wave packet would be small compared to $\lambda/2$ and if such packet would not spread during the interaction time τ, the results of the quantum-mechanical solution of the problem should differ significantly from those described in the preceding section and should correspond completely to the classical treatment at any value of β. The solution of the quantum problem of the Kapitza–Dirac effect in terms of electronic wave packets will be described separately.

4.2. Classical angular distribution function of scattered electrons

In a beam of classical electrons with a homogeneous distribution with respect to the initial coordinate z_0, the number of electrons having a value of the coordinate z in the interval from z_0 to z_0+dz_0 obviously equals

$$dN(z_0)=N_0\frac{dz_0}{\lambda/2}, \tag{37}$$

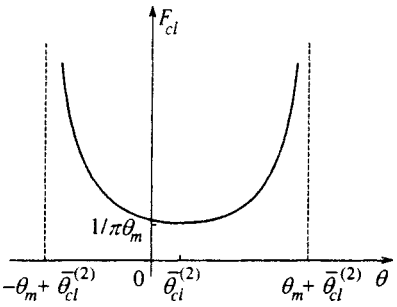

FIG. 5. Distribution function of scattered electrons found from the classical theory (40).

where N_0 is the total number of electrons in the periodicity interval $\lambda/2$. Before scattering (at $t=t_0$) all the electrons in the beam have the same velocity v_0 and the same glancing angle α. In view of the conservation of the number of particles, after scattering the same number of electrons (37) will have the velocity v, and their direction will correspond to the interval of angles $[\theta, \theta+d\theta]$:

$$dN(\theta)\cong N_0 F_{\rm cl}(\theta)d\theta=\frac{N_0 d\theta}{\lambda/2}\sum_i\left|\frac{dz_0^{(i)}(\theta)}{d\theta}\right|$$

$$=\frac{N_0 d\theta}{\lambda/2}\sum_i\frac{1}{|(d\theta(z_0)/dz_0)^{(i)}|}, \tag{38}$$

where $F_{\rm cl}(\theta)$ is the classical distribution function and $z_0(\theta)$ is a function, which is the inverse of $\theta(z_0)$. This function can be multivalued (see below), and, for this reason, the sum over i, where $i=1,2$ labels the solutions of the equation $\theta=\theta(z_0)$ with respect to z_0, appears in the definitions (38). The explicit form of $\theta(z_0)$ is given by the equation from first-order perturbation theory with respect to U_0 [Eq. (34)] with the addition of a correction term, which takes into account the contribution of the second order to the mean scattering angle (36):

$$\theta=\theta_m\sin(2kz_0-\varphi)+\overline{\theta_{\rm cl}^{(2)}}, \tag{39}$$

where θ_m and φ are defined by Eqs. (34). Using the definition (38) and Eq. (39) for $\theta(z_0)$, we ultimately find the distribution function $F_{\rm cl}(\theta)$ of a beam of classical electrons scattered on a standing wave:

$$F_{\rm cl}(\theta)=\frac{1}{\pi\sqrt{\theta_m^2-(\theta-\overline{\theta_{\rm cl}^{(2)}})^2}} \tag{40}$$

in the range of angles $-\theta_m+\overline{\theta_{\rm cl}^{(2)}}<\theta<\theta_m+\overline{\theta_{\rm cl}^{(2)}}$ and $F_{\rm cl}(\theta)=0$ outside that range. The function $F_{\rm cl}(\theta)$ (40) is depicted in Fig. 5. It clearly displays striking differences from the distribution function appearing in the model of plane-wave scattering (Fig. 4). Therefore, direct measurements of the distribution of scattered electrons can be a source of information regarding the applicability of a particular model. The distribution function (40) is normalized to unity. It gives the correct value of the mean scattering angle $\overline{\theta}=\overline{\theta_{\rm cl}^{(2)}}$ (36). The function $F_{\rm cl}(\theta)$ is asymmetric: it is shifted as a whole by $\overline{\theta_{\rm cl}^{(2)}}$ relative to $\theta=0$, and just this asymmetry causes the mean scattering angle to differ from zero. As the field strength is lowered, the width $2\theta_m$ of the localization interval $[-\theta_m+\overline{\theta_{\rm cl}^{(2)}}, \theta_m+\overline{\theta_{\rm cl}^{(2)}}]$ of the distribution function $F_{\rm cl}(\theta)$ (40) decreases proportionally to $U_0\propto\varepsilon_0^2$, while the degree of its asymmetry decreases more rapidly, in proportion to $U_0^2\propto\varepsilon_0^4$. By virtue of the normalization to unity it hence follows that in the limit $\varepsilon_0\to 0$ the function $F_{\rm cl}(\theta)$ transforms into $\delta(\theta)$, which corresponds to the absence of scattering in an infinitely weak field.

5. DISCUSSION OF RESULTS

Let us briefly formulate and discuss the main results obtained.

1. A quantum-mechanical analysis of the scattering of an electron on a standing light wave with assignment of the initial electronic wave function in the form of a plane wave has revealed two scattering regimes, which differ with respect to the value of the parameter β (19). The region $\beta\ll 1$ corresponds to the familiar picture of Bragg scattering: the scattering is effective only if the initial glancing angle α of the electron is close to the Bragg angle, i.e., if $\alpha=-\alpha_{\rm Br}$ or $\alpha=\alpha_{\rm Br}$, and the dependence of the mean electron scattering angle on α, $\overline{\theta}(\alpha)$, has the form of a curve with sharp peaks at $\mp\alpha_{\rm Br}$ (the solid curve in Fig. 2). Conversely, the classical limit of the quantum theory of scattering is realized in the region $\beta\gg 1$. In this case the sharp dependence of the scattering angle on the initial glancing angle of the electron vanishes. The dependence of $\overline{\theta}$ on α is characterized by a smooth broad curve (the solid curve in Fig. 3), whose maximum and minimum correspond to values of α which significantly exceed $\alpha_{\rm Br}$ in absolute value. In the limit of very large β, the result of the quantum-mechanical calculation in the model of plane-wave scattering ceases to depend on Planck's constant $\hbar$.

2. A classical calculation of the mean scattering angle of a beam of electrons uniformly distributed with respect to the initial transverse coordinate z_0 on a standing wave has been performed. The result obtained coincides with the classical limit of the quantum theory of the scattering of an electron, whose initial wave function is assigned by a plane wave. The classical treatment is not restricted by any assumptions regarding the value of the quantum parameter β. When $\beta\gg 1$, the results of the classical and quantum treatments coincide (Fig. 3). Conversely, when $\beta\ll 1$, these results differ significantly, as is clearly seen from Fig. 2.

3. Even more striking differences between the classical and quantum theories are displayed for the distribution function of scattered electrons found (Figs. 4 and 5). It should be concluded from the foregoing statements that direct experimental measurements of the angular distribution of scattered electrons can serve as a basis for determining whether the classical or quantum model of plane-wave scattering more faithfully describes a given electron beam.

According to the foregoing, when $\beta\ll 1$, conclusions regarding the applicability or inapplicability of a particular model can be drawn on the basis of measurements of the dependence of the mean scattering angle on the initial glancing angle $\overline{\theta}(\alpha)$. In the quantum model of plane-wave scat-

tering, narrow peaks should appear on the $\overline{\theta}(\alpha)$ curve at $\alpha = \mp\alpha_{\mathrm{Br}}$, and in the classical theory they should appear at considerably smaller values of α, i.e., at $\alpha = \mp\alpha_{\mathrm{cl}} = \mp\beta\alpha_{\mathrm{Br}}$ (22) (see Fig. 2), where $\beta\alpha_{\mathrm{Br}} \ll \alpha_{\mathrm{Br}}$.

It seems to us that these differences are caused specifically by the use of a plane wave as the initial electronic wave function in the existing version of the quantum theory. In an alternative version of the quantum theory the initial state of an electron is described in the form of a localized wave packet.[14] If the packet width is smaller than the inhomogeneity scale length $\lambda/2$ and if such a packet does not spread during the interaction time τ, complete agreement between the results of the classical and quantum descriptions should be expected. This program has not yet been fully implemented and calls for a separate treatment.

In terms of wave packets, the experiment proposed and described above for measuring the angular dependence of scattered electrons and the $\overline{\theta}(\alpha)$ curve can be regarded as a method for determining the wave packet width Δr. If the measurement results are close to the predictions of the classical theory, then $\Delta r \ll \lambda/2$. If the measurement results are close to what follows from the quantum theory of plane-wave scattering, they can be regarded as evidence that the packet width is very large, $\Delta r \gg \lambda/2$.

The wave packet width for beam electrons is a parameter which is usually not manifested in any way and is not discussed. Electrons in a beam are treated either as an ensemble of classical point objects (with a radius of the order of the classical electron radius $r_0 = e^2/mc^2 \approx 2.5\times 10^{-13}$ cm) or as a quantum plane wave. In this paper, first, attention has been focused on the fact that these two descriptions are not equivalent, and, second, an experiment has been proposed for drawing conclusions regarding the value of this cryptic parameter of beam electrons, i.e., the quantum wave packet width, which characterizes their state before scattering.

Finally, let us clarify the specific features of a standing wave as opposed to other objects on which electron scattering can occur. The inhomogeneity scale length in a standing wave $\lambda/2$ is of the order of a micron ($\sim 10^{-4}$ cm) in the optical wavelength range and is thus much greater than in the case of scattering on atomic targets. Just this circumstance makes it reasonable to pose the question of the scattering of localized wave packets with a width $\Delta r < \lambda/2$. If targets and wave packets have small dimensions, the wave packets ''spread'' very rapidly, and the condition $\Delta r < \lambda/2$ ceases to hold. For example, a wave packet of atomic size a ($\sim 10^{-8}$ cm) spreads during an atomic time period ($\sim 10^{-16}$ s). Therefore, when electrons are scattered on atoms, the inequality $\Delta r \gg a$ always holds, and the state of the incident electrons can be approximated by a plane wave. Conversely, wave packets of micron size do not spread for a fairly long time ($\sim 10^{-8}$ s), enabling the realization of the cases of both $\Delta r > \lambda/2$ and $\Delta r < \lambda/2$. In addition, at the present time there are fully realistic methods for creating such electronic wave packets in a strictly controlled manner. One of these methods is the multiphoton ionization of atoms by a laser field. In this case the size of the wave packets formed in the continuum can be determined, for example, by the duration of the pulse of ionizing radiation. The efficiency of such a scheme for forming localized wave packets was demonstrated in the experiment in Ref. 10, where scattering of the electrons obtained in such a way on the ponderomotive potential in the focus of a second laser was observed. Just such a scheme may prove to be very convenient for investigating the features of scattering on a standing wave described above.

*)E-mail: fedorov@theor.msk.ru

[1] P. L. Kapitza and P. A. M. Dirac, Proc. Cambridge Philos. Soc. **29**, 297 (1933).
[2] M. B. Fedorov, Zh. Éksp. Teor. Fiz. **52**, 1434 (1967) [Sov. Phys. JETP **25**, 952 (1967)].
[3] M. B. Fedorov, *Electrons in a Strong Light Field* [in Russian], Nauka, Moscow (1991), p. 45.
[4] L. S. Bartell, J. Appl. Phys. **38**, 1561 (1967).
[5] L. S. Bartell, R. R. Roskos, and H. B. Thompson, Phys. Rev. **166**, 1494 (1968).
[6] Y. Takeda and I. Matsui, J. Phys. Soc. Jpn. **15**, 1202 (1968).
[7] H. C. Pfeiffer, Phys. Lett. A **26**, 326 (1968).
[8] H. Schwarz, Phys. Lett. A **43**, 457 (1973).
[9] L. S. Bartell, Phys. Lett. A **27**, 236 (1968).
[10] P. H. Bucksbaum, M. Bashkansky, and T. J. McIlrath, Phys. Rev. Lett. **58**, 349 (1987).
[11] P. H. Bucksbaum, D. W. Schumacher, and M. Bashkansky, Phys. Rev. Lett. **61**, 1182 (1988).
[12] C. I. Moore, J. P. Knauer, and D. D. Meyerhofer, Phys. Rev. Lett. **74**, 2439 (1995).
[13] C. Kittel, *Introduction to Solid State Physics*, 2nd ed., Wiley, New York (1956) [Russ. transl., Fizmatgiz, Moscow (1963), Chap. 3].
[14] E. Schrödinger, Naturwissenschaften **28**, 664 (1926).
[15] M. V. Fedorov, S. P. Goreslavsky, and V. S. Letokhov, Phys. Rev. E **55**, 1015 (1997).
[16] V. G. Minogin, M. V. Fedorov, and V. S. Letokhov, Opt. Commun. **140**, 250 (1997).
[17] D. R. Bitouk and M. V. Fedorov, Opt. Express **2**, 404 (1998); Laser Phys. **8**, 544 (1998); Phys. Rev. A **58**, 1195 (1998).

Translated by P. Shelnitz

Dynamics of plasma bunches in slowly varying external fields

D. S. Dorozhkina*) and V. E. Semenov†)

Institute of Applied Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia
(Submitted 10 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 885–901 (September 1999)

We use the quasineutrality approximation and the method of moments to analyze a system of kinetic equations that describes the expansion into vacuum of a plasma bunch generally containing several species of charged particles. For a two-component collisionless plasma in slowly varying external potential fields, we obtain a complete description of the dynamics of the matrices of centered second moments of the particle velocity distribution functions. We construct a new class of self-similar solutions of the kinetic equations in which the moments of the distribution functions act as parameters. These solutions are found to be valid for any mass and energy ratios of the constituent particles and generally describe the dynamics of a plasma bunch that is asymmetric in space. For a symmetric bunch we also find an analytical solution corresponding to the presence of eddy electric currents in the plasma, while for an asymmetric bunch we find that interparticle collisions, which give rise to anisotropy in the process of expansion of plasma into vacuum, play an important role. The method developed in the paper is used to study the acceleration and compression of a plasma bunch in time-dependent magnetic fields with a mirror configuration.
[S1063-7761(99)01009-4]

1. INTRODUCTION

In the absence of external fields, the dynamics of plasma in vacuum is quite obvious. A plasma bunch that was initially localized in space expands without limit and cools off. But when an external magnetic field is switched on, this bunch can be compressed and accelerated as a whole. Such processes are of substantial interest to astrophysicists and researchers working in the field of controlled fusion. An important element in such investigations is the detailed study of the dynamics of the bunch in free space. Note that expansion into vacuum of a semibounded plasma, studied earlier by many researchers,[1–5] and the results of generalizing this one-dimensional problem to the three-dimensional case[6,7] actually do not provide a correct description of the free dynamics of a plasma bunch. The problem is that the models used in the above studies presupposed the existence of an unlimited supply of energy and particles in the plasma, which to a great extent corresponds to a situation in which continuously operating sources are present in a certain region in space. Free expansion of a bounded bunch of plasma into vacuum is accompanied by the plasma cooling off in space, a process that earlier was accounted for only in the phenomenological hydrodynamic approach.[8] However, recently real advances in this field of research have been achieved,[9–14] which finally made it possible to construct, for two-component plasmas, an analytical solution of the collisionless kinetic equations in the self-consistent field generated by charge separation.

In the present paper we analyze in detail these solutions and generalize some of the results to the case of a dense plasma in which interparticle collisions play an important role and to the case of a multicomponent plasma with particles of different species. The method we develop is used to construct a solution for the Vlasov kinetic equations for a plasma bunch in slowly varying external potential fields acting on the plasma particles. The results are used to study the acceleration and compression of a plasma bunch in time-dependent magnetic fields with a mirror configuration.

2. STATEMENT OF THE PROBLEM: THE STARTING EQUATIONS

The dynamics of a plasma bunch in external potential fields is described in the general case by a system of kinetic equations for the velocity distribution of the particles of each species, $f_\alpha(\mathbf{v},\mathbf{r},t)$:

$$\frac{\partial f_\alpha}{\partial t}+(\mathbf{v}\cdot\boldsymbol{\nabla}_\mathbf{r})f_\alpha-\frac{Z_\alpha e}{m_\alpha}(\boldsymbol{\nabla}_\mathbf{r}\varphi\cdot\boldsymbol{\nabla}_\mathbf{v})f_\alpha - \frac{1}{m_\alpha}(\boldsymbol{\nabla}_\mathbf{r}U_\alpha\cdot\boldsymbol{\nabla}_\mathbf{v})f_\alpha=I_\alpha,$$
$$\boldsymbol{\nabla}_\mathbf{r}\equiv\sum_{k=1}^{3}\mathbf{e}_k\frac{\partial}{\partial r_k},\quad \boldsymbol{\nabla}_\mathbf{v}\equiv\sum_{k=1}^{3}\mathbf{e}_k\frac{\partial}{\partial v_k}, \tag{1}$$

where e is the absolute value of the elementary charge, Z_α and m_α are the charge number and mass of a particle of the α species, $\varphi(\mathbf{r},t)$ is the potential of the electric field generated in the course of plasma expansion due to charge separation, $U_\alpha(\mathbf{r},t)$ is the potential of the external field acting on the particles of the α species, I_α is the respective collision integral, r_k and v_k are the components of the radius vector $\mathbf{r}$ and the velocity vector $\mathbf{v}$, and $\mathbf{e}_k$ is the unit vector whose direction coincides with that of r_k. Below we limit ourselves to elastic collisions between the plasma particles, assuming all along that the collision integral does change the density of

particles of each species, the momentum density in the plasma, and the density of the plasma kinetic energy, i.e.,

$$\int I_\alpha \, d\mathbf{v} = 0, \tag{2}$$

$$\sum_\alpha m_\alpha \int v_k I_\alpha d\mathbf{v} = 0, \quad k = 1,2,3, \tag{3}$$

$$\sum_\alpha \sum_{k=1}^{3} \int m_\alpha v_k^2 I_\alpha d\mathbf{v} = 0. \tag{4}$$

In analyzing the expansion of a dense plasma we can find the potential φ of the field generated in the course of plasma expansion due to charge separation from the quasineutrality condition

$$\sum_\alpha Z_\alpha n_\alpha = 0, \quad n_\alpha(\mathbf{r},t) \equiv \int f_\alpha(\mathbf{v},\mathbf{r},t) d\mathbf{v}, \tag{5}$$

where $n_\alpha(\mathbf{r},t)$ is the density of particles of species α. In the simplest case, when there are no eddy currents in the plasma, Eq. (5) makes it possible to derive an explicit expression for the strength of the ambipolar electric field in terms of the particle distribution functions:

$$e \frac{\partial \varphi}{\partial r_j} = - \frac{\sum_\alpha \sum_{k=1}^{3} \frac{\partial}{\partial r_k} \int v_k v_j Z_\alpha f_\alpha d\mathbf{v} + \sum_\alpha \frac{Z_\alpha}{m_\alpha} \frac{\partial U_\alpha}{\partial r_j} \int f_\alpha(\mathbf{v},\mathbf{r},t) d\mathbf{v}}{\sum_\alpha \frac{Z_\alpha^2}{m_\alpha} \int f_\alpha(\mathbf{v},\mathbf{r},t) d\mathbf{v}}. \tag{6}$$

Plugging (6) into the initial kinetic equations, we arrive at a system of integro-differential equations. Below we seek the solutions to these equations with fixed velocity distributions of the particles of each species.

Using the quasineutrality approximation (5), we derived in Ref. 14 a self-similar solution of collisionless ($I_\alpha = 0$) Vlasov kinetic equations (1) describing the free ($U_\alpha = 0$) expansion into vacuum of a two-component plasma bunch with arbitrary ratios of particle masses and energies of random motion. Here the laws we found that govern the expansion process are of a universal nature and can be derived for arbitrary initial velocity distribution functions, including those that do not correspond to a self-similar solution but agree with the quasineutrality condition (5). We believe that this fact plays an important role in the analysis of more general cases of plasma expansion into vacuum (i.e., cases that allow for interparticle collisions) and of the dynamics of a plasma bunch in external fields. Hence we begin with the results of a general study by the method of moments of Eqs. (1), (5), and (6), which will then be used to obtain exact solutions.

3. THE METHOD OF MOMENTS

We define the operation of averaging of an arbitrary function $\Psi(\mathbf{v},\mathbf{r},t)$ over the distribution function f_α as the calculation of the functional

$$\langle \Psi \rangle_\alpha \equiv \frac{1}{N_\alpha} \int \int \Psi(\mathbf{v},\mathbf{r},t) f_\alpha(\mathbf{v},\mathbf{r},t) d\mathbf{v}\, d\mathbf{r}, \tag{7}$$

where $N_\alpha \equiv \int\int f_\alpha(\mathbf{v},\mathbf{r},t) d\mathbf{v} d\mathbf{r}$ is the conserved number of particles of the α species. Then, because of the quasineutrality of the plasma [the condition (5)], the following relationships hold for every function $\tilde{\Psi}(\mathbf{r},t)$ that does not explicitly depend on the particle velocity $\mathbf{v}$:

$$\sum_\alpha Z_\alpha N_\alpha \langle \tilde{\Psi}(\mathbf{r},t) \rangle_\alpha = 0,$$

$$\sum_\alpha Z_\alpha N_\alpha \sum_{k=1}^{3} \left\langle v_k \frac{\partial \tilde{\Psi}}{\partial r_k} \right\rangle_\alpha = \sum_\alpha Z_\alpha N_\alpha \left[\frac{d}{dt} \langle \tilde{\Psi} \rangle_\alpha - \left\langle \frac{\partial}{\partial t} \tilde{\Psi} \right\rangle_\alpha \right] = 0. \tag{8}$$

On the other hand, for the elastic–collision model (2)–(4) we can easily find that

$$\int \int \tilde{\Psi}(\mathbf{r},t) I_\alpha d\mathbf{v}\, d\mathbf{r} = 0,$$

$$\sum_\alpha M_\alpha \int \int v_k \tilde{\Psi} I_\alpha d\mathbf{v}\, d\mathbf{r} = 0, \quad k = 1,2,3, \tag{9}$$

$$\sum_\alpha M_\alpha \sum_{k=1}^{3} \int \int v_k^2 \tilde{\Psi}(\mathbf{r},t) I_\alpha d\mathbf{v}\, d\mathbf{r} = 0,$$

where $M_\alpha \equiv m_\alpha N_\alpha$ is the total mass of the α component of the plasma.

In the absence of external fields ($U_\alpha = 0$), by combining (8) and (9) with the kinetic equations (1) we can derive differential relationships that link the various moments of the particle distribution functions:

$$\frac{d\langle r_k \rangle_\alpha}{dt} = \langle v_k \rangle_\alpha, \tag{10}$$

$$\sum_\alpha M_\alpha \frac{d\langle v_k \rangle_\alpha}{dt} \equiv M \frac{d^2}{dt^2} R_k = 0, \tag{11}$$

$$\frac{d\langle \tilde{r}_k \tilde{r}_j \rangle_\alpha}{dt} = \langle \tilde{r}_k \tilde{v}_j \rangle_\alpha + \langle \tilde{v}_k \tilde{r}_j \rangle_\alpha, \tag{12}$$

$$\frac{d}{dt} \sum_\alpha M_\alpha \langle \tilde{r}_k \tilde{v}_j \rangle_\alpha = \sum_\alpha M_\alpha \langle \tilde{v}_k \tilde{r}_j \rangle_\alpha, \tag{13}$$

$$\frac{d}{dt}\sum_{\alpha}\sum_{k=1}^{3} M_{\alpha}\langle\tilde{v}_k^2\rangle_{\alpha}=0, \tag{14}$$

where $M\equiv\Sigma_{\alpha}M_{\alpha}$ is the total mass of the plasma, the $R_k(t)=\Sigma_{\alpha}\mu_{\alpha}\langle r_k\rangle_{\alpha}$ are the coordinates of the center of mass of the plasma bunch, with $\mu_{\alpha}\equiv M_{\alpha}/M$, and the $\tilde{r}_k\equiv r_k-R_k$ and $\tilde{v}_k\equiv v_k-dR_k/dt$ are the components of the radius vector $\tilde{\mathbf{r}}$ and the velocity vector $\tilde{\mathbf{v}}$ in the reference frame in which the center of mass of the bunch is at rest.

In the general case of a multicomponent plasma, the system of equations (10)–(14) for the moments may be open, Nevertheless, such a system possesses a number of nontrivial integrals, which makes it possible to determine the evolution of the characteristic size of the plasma bunch and the energy of the thermal (chaotic) motion of the constituent particles. Indeed, let us define the characteristic size l of the plasma in terms of centered second moments:

$$l^2(t)\equiv\sum_{\alpha}\sum_{k=1}^{3}\mu_{\alpha}\langle\tilde{r}_k^2\rangle_{\alpha}. \tag{15}$$

Then from (12)–(14) it follows that the dynamics of $l^2(t)$ (i.e., the expansion of the plasma) is completely determined by the initial values of the second moments of the distribution functions f_{α}:

$$\frac{d^2l^2}{dt^2}=2\sum_{\alpha}\sum_{k=1}^{3}\mu_{\alpha}\langle\tilde{v}_k^2\rangle_{\alpha}\equiv 4\frac{W}{M}=\text{const}, \tag{16}$$

$$\frac{dl^2}{dt}=2\sum_{\alpha}\sum_{k=1}^{3}\mu_{\alpha}\langle\tilde{r}_k\tilde{v}_k\rangle_{\alpha}, \tag{17}$$

where W is the total kinetic energy of motion of the plasma particles in the center-of-mass reference frame.

The cooling of the plasma during expansion is described by formulas that follow directly from Eqs. (12) and (13):

$$\sum_{\alpha}\sum_{k=1}^{3}M_{\alpha}\left\langle\left(\tilde{v}_k-\frac{\tilde{r}_k}{l}\frac{dl}{dt}\right)^2\right\rangle_{\alpha}=2W-M\left(\frac{dl}{dt}\right)^2. \tag{18}$$

Here the right-hand side decreases with time in inverse proportion to l^2, which follows from the first integral of Eq. (16),

$$\left[\frac{2W}{M}-\left(\frac{dl}{dt}\right)^2\right]l^2(t)=\text{const}. \tag{19}$$

Thus, the thermal velocity spread of particles of each species in relation to the ''hydrodynamic'' velocity

$$\mathbf{u}=\frac{d\mathbf{R}}{dt}+\frac{\mathbf{r}-\mathbf{R}}{l}\frac{dl}{dt},$$

where $\mathbf{R}$ is the vector with components R_k, decreases in the course of the expansion in inverse proportion to the size l of the plasma bunch. Accordingly, we can state that after a long time has passed [and the characteristic scale $l(t)$ becomes much larger than the initial scale $l(0)$], the expansion process sets into a self-similar regime with the hydrodynamic velocity $\mathbf{u}$. In this regime all spatial scales for each plasma component increase in proportion to l:

$$n_{\alpha}(\mathbf{r},t)=\frac{l(0)}{l(t)}\tilde{N}_{\alpha}\left(\frac{\mathbf{r}-\mathbf{R}}{l(t)}\right). \tag{20}$$

Unfortunately, in the general case of a multicomponent plasma, Eqs. (12)–(14) do not describe the variation of the second spatial moments $\langle\tilde{r}_k\tilde{r}_j\rangle_{\alpha}$, i.e., the evolution of the shape of the plasma bunch and the relative distribution in this bunch of particles of different species remain undefined. However, if the plasma has only two components, it is obvious that in view of the quasineutrality condition (5) the matrices of the second spatial moments are the same for the two species of the oppositely charged particles. Here Eqs. (16)–(19) provide a fairly complete description of the expansion into vacuum of, say, a spherically symmetric plasma bunch. In the simplest case of expansion of a collisionless two-component plasma in the absence of electric currents, i.e., when

$$\sum_{\alpha=1,2}Z_{\alpha}\int v_k f_{\alpha}\,d\mathbf{v}=0.$$

the matrices of the mixed second moments $\langle\tilde{r}_k\tilde{v}_j\rangle_{\alpha}$ are the same for the two particle species. Hence the system of equations for the second moments is found to be closed:[1)]

$$\frac{d\langle\tilde{r}_k\tilde{r}_j\rangle}{dt}=\langle\tilde{r}_k\tilde{v}_j\rangle+\langle\tilde{v}_k\tilde{r}_j\rangle, \tag{21}$$

$$\frac{d\langle\tilde{r}_k\tilde{v}_j\rangle}{dt}=\sum_{\alpha}\mu_{\alpha}\langle\tilde{v}_k\tilde{v}_j\rangle_{\alpha}, \tag{22}$$

$$\frac{d}{dt}\sum_{\alpha}\mu_{\alpha}\langle\tilde{v}_k\tilde{v}_j\rangle_{\alpha}=0. \tag{23}$$

As a result, plasma expansions along the different coordinate axis are independent of each other, i.e., for the scales $l_k(t)$ ($l_k^2(t)\equiv\langle\tilde{r}_k^2\rangle$) we can write equations similar to Eqs. (16) and (17):

$$\frac{d^2l_k^2}{dt^2}=2\sum_{\alpha}\mu_{\alpha}\langle\tilde{v}_k^2\rangle_{\alpha}\equiv 4\frac{W_k}{M}=\text{const}, \tag{24}$$

$$\frac{dl_k^2}{dt}=2\langle\tilde{r}_k\tilde{v}_k\rangle, \tag{25}$$

where W_k is the kinetic energy of motion of the plasma particles along the direction of $\tilde{r}_k$ in the center-of-mass reference frame. Accordingly, the cooling of the plasma, i.e., the decrease in the energy of chaotic motion of the particles, along one coordinate axis is independent of the cooling along the other axes:

$$l_k^2(t)\sum_{\alpha}\mu_{\alpha}\left\langle\left(\tilde{v}_k-\frac{\tilde{r}_k}{l_k}\frac{dl_k}{dt}\right)^2\right\rangle_{\alpha}\equiv l_k^2(t)\sum_{\alpha}\mu_{\alpha}V_{k\alpha}^2(t)=\text{const}. \tag{26}$$

Note that in this simple example of a two-component collisionless ''current-free'' plasma, the method of moments can also be successfully employed in the study of the dynamics of a plasma bunch in slowly varying external potential fields. As long as the bunch size is small compared to the characteristic gradient lengths of the U_{α}, the later can be

represented in the form of power series in the spatial variables $\tilde{r}_k$ in the neighborhood of the center of mass of the plasma bunch, $\mathbf{r}=\mathbf{R}$. Keeping only terms whose order is no higher than the second in $\tilde{r}_k$, we can use (1) and (5) to obtain equations [that replace Eqs. (11)–(14)] for the moments of the particle-velocity distribution functions:

$$\frac{d^2R_k}{dt^2}=a_k, \tag{27}$$

$$\frac{d\langle\tilde{r}_k\tilde{r}_j\rangle}{dt}=\langle\tilde{r}_k\tilde{v}_j\rangle+\langle\tilde{v}_k\tilde{r}_j\rangle, \tag{28}$$

$$\frac{d\langle\tilde{r}_k\tilde{v}_j\rangle}{dt}=\sum_\alpha \mu_\alpha\langle\tilde{v}_k\tilde{v}_j\rangle_\alpha-\sum_{i=1}^{3} b_{ij}\langle r_i\tilde{r}_k\rangle, \tag{29}$$

$$\frac{d}{dt}\sum_\alpha \mu_\alpha\langle\tilde{v}_k\tilde{v}_j\rangle_\alpha=-\sum_{i=1}^{3}\{b_{ij}\langle\tilde{r}_i\tilde{v}_k\rangle+b_{ik}\langle\tilde{r}_i\tilde{v}_j\rangle\}. \tag{30}$$

Here the components $a_k(t)$ of the acceleration vector and the matrix elements $b_{kj}(t)$ can be expressed in terms of the first and second derivatives, respectively, of the effective potential $U(\mathbf{r},t)\equiv\Sigma_\alpha N_\alpha U_\alpha(\mathbf{r},t)$ with respect to the coordinates at the point coinciding with the center of mass of the plasma bunch:

$$a_k\equiv-\frac{1}{M}\frac{\partial U}{\partial r_k}\bigg|_{r_k=R_k}, \tag{31}$$

$$b_{kj}\equiv\frac{1}{M}\frac{\partial^2 U}{\partial r_k\partial r_j}\bigg|_{r_k=R_k,\,r_j=R_j}. \tag{32}$$

Equations (27)–(30) clearly show that the gradient of the effective potential determines the law of motion of the center of mass of the plasma bunch and has no effect on the dynamics of the characteristic scales of the plasma. The spatial structure of the bunch depends only on the second derivatives of the external field U. Hence the acceleration of the plasma bunch and the expansion of the bunch can be controlled independently. In particular, if initially all the matrices of the second moments of the plasma-particle velocity distribution functions are diagonal in a certain reference frame and, at the same time, for $k\neq j$ the second derivatives of the effective potential are zero ($b_{kj}=0$), the expansion of the bunch along one axis is independent of that along the other axes and is described by the system of equations

$$\frac{d^2l_k^2}{dt^2}+2b_{kk}(t)l_k^2(t)=4\frac{W_k}{M}, \tag{33}$$

$$\frac{d}{dt}\frac{W_k}{M}=-\frac{1}{2}b_{kk}(t)\frac{dl_k^2}{dt}. \tag{34}$$

This system of equations has a first integral similar to (19):

$$\left[\frac{2W_k}{M}-\left(\frac{dl_k}{dt}\right)^2\right]l_k^2(t)\equiv\sum_\alpha \mu_\alpha[V_{k\alpha}(t)l_k(t)]^2=\text{const}, \tag{35}$$

where by analogy with (26) we have used $V_{k\alpha}(t)$ to denote the mean-square velocity of the thermal motion of the particles of the α species along the direction specified by $\tilde{r}_k$. Thus the system of equations consisting of (33) and (34) reduces to a single second-order equation:

$$l_k^3(t)\frac{d^2l_k}{dt^2}+b_{kk}(t)l_k^4(t)=\sum_\alpha \mu_\alpha[V_{k\alpha}(t)l_k(t)]^2=\text{const}. \tag{36}$$

According to this equation, when the second derivatives of the external-field potential are positive ($b_{kk}>0$), there is a limit to the bunch expansion. For $b_{kk}l_k^2>\Sigma_\alpha\mu_\alpha V_{k\alpha}^2$ the characteristic plasma scales may even decrease.

Note that the condition for localization of a plasma bunch ($b_{kk}>0$) coincides with the condition for stable acceleration of the plasma particles in the traveling wave of the effective potential $U(\mathbf{r},t)$. Hence, using the traveling wave of the external field, we can ensure the acceleration of the plasma bunch in the ''self-focusing'' regime,'' with the bunch remaining localized in space.

Thus, the method of moments can be effectively used to solve problems of the dynamics of a plasma bunch in slowly varying external potential fields. In the simplest case of a collisionless two-component ''current-free'' plasma, the method yields a closed system of equations for the elements $\langle\tilde{r}_k\tilde{r}_j\rangle$ of the matrix of centered second moments, and the analysis of the solutions of this system for arbitrary initial conditions does not present serious difficulties. On the other hand, it is possible to obtain a solution of the kinetic equations in this case (the specific dependence of the f_α on their arguments) only for a much narrower class of initial conditions corresponding to the self-similar nature of the plasma-bunch parameters. Nevertheless, it would be interesting to find such a solution, since it would provide detailed information about the evolution of the plasma.

4. SELF-SIMILAR SOLUTIONS OF THE KINETIC EQUATIONS FOR A TWO-COMPONENT PLASMA

In this section we discuss examples of analytically solutions of the system of two kinetic equations in the quasineutral approximation. Such solutions are obtained under conditions corresponding to the above case of a two-component collisionless ''current-free'' plasma (Sec. 4.1) and in the presence of eddy electric currents in the plasma (Sec. 4.2). In addition, in Sec. 4.3 we study the example of expansion into vacuum of a bunch of dense plasma, in which interparticle collisions (or other processes responsible for isotropizing the particle distribution in the velocity space) play an important role.

4.1. Collisionless ''current-free'' plasma

The determination of the analytical solutions of the kinetic problem of the dynamics of a two-component collisionless plasma bunch in external fields is based on the methods developed in Refs. 11–14, where the self-similar expansion of a plasma bunch in vacuum was studied ($U_\alpha=0$). It was found that in slowly varying external potential fields when the matrix b_{kj} of the second derivatives of the external-field potential is diagonal in the reference frame in which the center of mass of the plasma bunch is at rest, the form of the

self-similar solutions of the kinetic equations coincides with the exact solution in free space. All differences are due solely to the nature of the time dependence of the first and second moments, which enter into the f_α as parameters:

$$f_\alpha(\mathbf{v},\mathbf{r},t)=\lambda_\alpha F(G_1^{(\alpha)},G_2^{(\alpha)},G_3^{(\alpha)}),\quad \alpha=1,2,$$

$$G_k^{(\alpha)}=\left(\frac{\tilde r_k}{l_k(t)}\right)^2+\left(\frac{\tilde v_k-\tilde u_k(\tilde r_k,t)}{V_{k\alpha}(t)}\right)^2,\quad k=1,2,3, \tag{37}$$

$$\tilde r_k\equiv r_k-R_k(t),\quad \tilde v_k\equiv v_k-\frac{dR_k}{dt},$$

where F is an arbitrary function of its arguments, the λ_α are normalization constants, the $R_k(t)$ are the coordinates of the center of mass of the plasma bunch in the laboratory reference frame, which vary according to (27), the $l_k(t)\equiv\sqrt{\langle\tilde r_k^2\rangle}$ are the scales of spatial localization of the plasma bunch, the $\tilde u_k(\tilde r_k,t)=\tilde r_k\langle\tilde r_k\tilde v_k\rangle/\langle\tilde r_k^2\rangle$ are the components of the average (hydrodynamic) particle velocity $\tilde{\mathbf{u}}(\mathbf{r},t)$ in the plasma-bunch center-of-mass reference frame, and the quantities $V_{k\alpha}^2(t)\equiv\langle(\tilde v_k-\tilde u_k)^2\rangle_\alpha$ describe the thermal spread of the particles of the α species in their velocities along each direction specified by $\tilde r_k$. Here the dynamics of all the moments in (37) is determined by Eqs. (36)[2)] and Eq. (25) from Sec. 3, and in the integral (35) each of the two terms separately is a constant

$$V_{k\alpha}(t)l_k(t)=\text{const}. \tag{38}$$

The solutions thus constructed are valid for any mass and initial-kinetic-energy ratios of the particles of the different species and describe the dynamics of a collisionless plasma bunch characterized by an arbitrary initial velocity distribution and, in general, by an anisotropic distribution of the plasma density in space.

Our results show that the average velocities $V_{k\alpha}$ of the thermal motion of particles of the different species change according to the adiabatic law (38), i.e., are inversely proportional to the corresponding size of the plasma bunch. The hydrodynamic velocity $\tilde{\mathbf{u}}(\mathbf{r},t)$ is the same for both components.

As the plasma freely expands into vacuum ($U_\alpha=0$), the thermal energy of the particles gradually transforms into the energy of their collective motion. Depending on the ratio of the initial velocities $V_{k\alpha}$, the ambipolar electric field accelerates particles with positive charge or with negative charge. More precisely, the average kinetic energy of the particles that initially had a lower thermal velocity increases. Accordingly, the average kinetic energy of the particles belonging to the other fraction decreases. The characteristic time of energy exchange between the plasma components is determined by the time it takes sound to propagate (with a speed equal to $\sqrt{2W_k/M}$) over distances of order the initial scale $l_k(0)$ of localization of the density of the plasma bunch along the respective axis.

A good way to illustrate the energy transfer is to write the integral velocity distribution of the particles of species α, $\phi_\alpha(\tilde{\mathbf{v}},t)=\int f_\alpha(\tilde{\mathbf{v}},\tilde{\mathbf{r}},t)\,d\tilde{\mathbf{r}}$, which at each moment in time remain similar to the initial distributions:

$$\phi_\alpha(\hat{\mathbf{v}},t)=\frac{1}{|Z_\alpha|}\phi_0\left(\frac{M\tilde v_x^2}{2W_{x\alpha}},\frac{M\tilde v_y^2}{2W_{y\alpha}},\frac{M\tilde v_z^2}{2W_{z\alpha}}\right)\times\prod_{k=1}^{3}\sqrt{\frac{M}{2W_{k\alpha}}}, \tag{39}$$

$$\frac{2W_{k\alpha}}{M}=[V_{k\alpha}(t)]^2+[\dot l_k(t)]^2, \tag{40}$$

where the function ϕ_0 is determined by the initial velocity distribution of the particles. Equation (39) states that in the case of self-similar expansion of a two-component plasma into vacuum the energy spectrum is conserved.

Returning to the solution (37) of the problem of the dynamics of the plasma bunch in external fields U_α, we note that in fact it cannot provide an absolutely exact description of the process we are interested in. The thing is that one consequence of the self-similar nature of the plasma motion is the quadratic dependence of the potential of the ambipolar electric field, φ, on the spatial coordinates $\tilde r_k$, which corresponds to the presence in space of a homogeneous electric charge density:

$$e\varphi(\tilde{\mathbf{r}},t)=\left(\sum_\alpha\frac{1}{M_\alpha}\right)^{-1}\sum_\alpha\frac{1}{Z_\alpha N_\alpha}\times\left[\sum_{k=1}^{3}\frac{V_{k\alpha}^2(0)l_k^2(0)}{2l_k^4(t)}\tilde r_k^2-\frac{1}{m_\alpha}U_\alpha\right]. \tag{41}$$

This means that generally the adopted model of the quasineutral dynamics of the plasma bunch is meaningless in regions where the plasma density is low. Nevertheless, for a dense plasma in which the frequencies $\omega_{p\alpha}\equiv 4\pi Z_\alpha^2e^2n_\alpha/m_\alpha$ of Langmuir oscillations satisfy the inequality

$$\sum_\alpha\omega_{p\alpha}^2\gg\sum_\alpha\frac{Z_\alpha}{|Z_\alpha|}\sum_{k=1}^{3}\left[\frac{V_{k\alpha}^2}{l_k^2(t)}-\frac{1}{m_\alpha}\frac{\partial^2U_\alpha}{\partial r_k^2}\right], \tag{42}$$

the quasineutrality condition (5) is violated only at the periphery of the plasma structure, far from the region where the bulk of the plasma is localized. In the presence of fields accelerating the plasma, the validity of the quasineutrality approximation requires that one more condition be met [in addition to (42)]: the electric field strength at the center of mass of the bunch must be small compared to the characteristic value of the field generated by one of the plasma components separately. In general, such a condition places an upper bound on the possible acceleration of the bunch.

Note that within our solution in the case of free expansion of a plasma into vacuum ($U_\alpha=0$), the right-hand side of (42) decreases as a function of time faster than the plasma density (because of the cooling of the plasma). Hence, with the passage of time the quasineutral description of the expansion of the plasma bunch remains valid in the plasma localization region.

4.2. Collisionless plasma with eddy electric currents

The above class of self-similar solution of the problem of expansion of a two-component plasma bunch into vacuum

corresponds to the case in which there are no currents in the plasma. Such a restriction guarantees that the system of equations (12)–(14) in the second moments of the distribution functions is closed. However, as noted in Sec. 3, this requirement is not needed if the plasma possesses a certain space symmetry. In particular, for a spherically symmetric plasma bunch the second spatial moments $\langle \tilde{r}_k \tilde{r}_j \rangle_\alpha$ can be written

$$\langle \tilde{r}_k \tilde{r}_j \rangle_\alpha = \frac{1}{3} l^2 \delta_{kj}, \quad \delta_{kj} = \begin{cases} 1, & k=j, \\ 0, & k \neq j, \end{cases} \tag{43}$$

where l^2 is determined by Eqs. (16) and (17) irrespective of whether there are eddy currents in the plasma. Here the off-diagonal matrix elements $\langle \tilde{r}_k \tilde{v}_j \rangle_\alpha$ of the mixed second moments, which generally do not coincide for particles of different species, are constant in time in view of conservation of the angular momentum of each component of the plasma:

$$\langle \tilde{r}_k \tilde{v}_j \rangle_\alpha + \langle \tilde{r}_j \tilde{v}_k \rangle_\alpha = \frac{1}{3} \delta_{kj} \frac{dl^2}{dt}, \tag{44}$$

$$\langle \tilde{r}_k \tilde{v}_j \rangle_\alpha - \langle \tilde{r}_j \tilde{v}_k \rangle_\alpha = \text{const}. \tag{45}$$

The symmetry also makes it possible to drop the requirement that there must be no eddy currents in the plasma when we construct analytical solutions of the kinetic equations. For instance, the functions

$$f_\alpha(\mathbf{v},\mathbf{r},t) = F_\alpha(G_r^{(\alpha)}, J_{12}, J_{23}, J_{31}), \tag{46}$$

$$G_r^{(\alpha)} = \frac{\tilde{\mathbf{r}}^2}{l^2(t)} + \frac{(\tilde{\mathbf{v}} - \tilde{\mathbf{w}})^2}{V_\alpha^2(t)}, \quad J_{kj} \equiv -J_{jk} = \tilde{r}_k \tilde{v}_j - \tilde{r}_j \tilde{v}_k,$$

are solutions of the Vlasov collisionless kinetic equations for arbitrary functions F_α if $l(t)$ satisfies Eq. (16),

$$V_\alpha(t) l(t) = \text{const}, \quad \tilde{\mathbf{w}} = \tilde{\mathbf{r}} \frac{1}{l(t)} \frac{dl}{dt},$$

and the potential of the ambipolar electric field is determined by the formula

$$e\varphi(\tilde{\mathbf{r}},t) = \left(\sum_\alpha \frac{1}{M_\alpha} \right)^{-1} \left(\sum_\alpha \frac{V_\alpha^2(t)}{Z_\alpha N_\alpha} \right) \frac{\tilde{\mathbf{r}}^2}{2l^2(t)}. \tag{47}$$

Here the quasineutrality approximation (5) partially limits the ambiguity in selecting the functions F_α:

$$\sum_\alpha \int Z_\alpha F_\alpha(\mathbf{v},\mathbf{r},t) d\mathbf{v} = 0. \tag{48}$$

The presence of the J_{kj} in the solution (46) makes it possible to describe the expansion of a plasma bunch with eddy currents. Note that although φ is spherically symmetric, the spatial distribution of the plasma density may be asymmetric.

Another example of free expansion into vacuum of a plasma with an electric current is the case of an axisymmetric plasma bunch, where the plasma density distribution and the potential φ of the ambipolar electric field depend only on two spatial variables, the longitudinal coordinate $\tilde{z}$ along the symmetry axis and the distance $\rho \equiv \sqrt{\tilde{x}^2 + \tilde{y}^2}$ from the axis. Here the analytical solution in the center-of-mass reference frame is

$$f_\alpha(\mathbf{v},\mathbf{r},t) = \lambda_\alpha F_\alpha(G_\rho^{(\alpha)}, J_z, G_z^{(\alpha)}),$$

$$G_\rho^{(\alpha)} = \frac{\rho^2}{l_\rho^2(t)} + \left(\frac{\tilde{v}_x - \tilde{x}\dot{l}_\rho / l_\rho}{V_{\rho\alpha}(t)} \right)^2 + \left(\frac{\tilde{v}_y - \tilde{y}\dot{l}_\rho / l_\rho}{V_{\rho\alpha}(t)} \right)^2,$$

$$\dot{l}_\rho \equiv \frac{dl_\rho}{dt}, \tag{49}$$

$$J_z = \tilde{x}\tilde{v}_y - \tilde{y}\tilde{v}_x, \quad G_z^{(\alpha)} = \left(\frac{\tilde{z}}{l_z(t)} \right)^2 + \left(\frac{\tilde{v}_z - \tilde{z}\dot{l}_z / l_z}{V_{z\alpha}(t)} \right)^2,$$

$$\dot{l}_z \equiv \frac{dl_z}{dt}, \quad V_{\rho\alpha}(t) l_\rho(t) = \text{const}, \quad V_z(t) l_z(t) = \text{const},$$

where $l_z(t) \equiv \langle \tilde{z}^2 \rangle$ and $l_\rho^2 \equiv \langle \rho^2 \rangle$ are the corresponding dimensions of the plasma bunch, which satisfy Eq. (24); $V_{z\alpha}(t)$ and $V_{\rho\alpha}(t)$ are the average thermal velocities of particles of the α species, defined as in (37); and the F_α are arbitrary functions, which, by virtue of quasineutrality, must meet the integral condition (48). The potential distribution is given by the formula

$$e\varphi(\tilde{\mathbf{r}},t) = \left(\sum_\alpha \frac{1}{M_\alpha} \right)^{-1} \sum_\alpha \frac{1}{Z_\alpha N_\alpha} \left[\frac{V_{z\alpha}^2(t)}{2l_z^2(t)} \tilde{z}^2 + \frac{V_{\rho\alpha}^2(t)}{2l_\rho^2(t)} \rho^2 \right]. \tag{50}$$

4.3. Effect of processes responsible for isotropization of the particle velocity distribution function on the expansion of a plasma bunch into vacuum

The solution of the problem of expansion into vacuum ($U_\alpha = 0$) of a two-component collisionless ''current-free'' plasma, Eq. (24), shows that when the initial velocity distribution of the particles is symmetric (when $W_k = W_j$ and $V_{k\alpha} = V_{j\alpha}$ for $k \neq j$), the possible initial anisotropy in the spatial distribution of the plasma ($l_k(0) \neq l_j(0)$ for $k \neq j$) disappears as the plasma bunch expands: $l_k(t)/l_j(t) \to 1$ as $t \to \infty$. At the same time, Eq. (38) suggests that the particle velocity distribution becomes locally anisotropic with the passage of time, so that $V_{k\alpha}(t)/V_{j\alpha}(t) \to l_k(0)/l_j(0)$ as $t \to \infty$. This is possible only if there are no collisions or other processes[3)] that facilitate the restoration of local isotropy of the distribution functions. The inverse of the process of expansion of dense plasma into vacuum is when the interparticle collisions maintain $V_{k\alpha}(t) = V_{j\alpha}(t)$ at each moment in time (for all values of k and j). In this limit, assuming that

$$V_{k\alpha}^2(t) = \frac{1}{3} \sum_{j=1}^{3} V_{j\alpha}^2(t) \equiv \frac{1}{3} V_\alpha^2(t), \tag{51}$$

we can close the system of equations (12)–(14) in the moments $\langle \tilde{r}_k^2 \rangle \equiv l_k^2$:

$$\frac{d^2 l_k^2}{dt^2} = 2 \left[\sum_\alpha \frac{\mu_\alpha V_\alpha^2(t)}{3} + \left(\frac{dl_k}{dt} \right)^2 \right], \tag{52}$$

$$\sum_\alpha \mu_\alpha V_\alpha^2(t) + \sum_{k=1}^{3} \left(\frac{dl_k}{dt} \right)^2 = \frac{2W}{M} = \text{const}. \tag{53}$$

Integration of Eqs. (52) and (53) yields the adiabatic law of plasma cooling:

$$\left[\prod_{k=1}^{3} l_k\right]^{2/3} \sum_{\alpha} \mu_\alpha V_\alpha^2(t) = \text{const}. \tag{54}$$

Here the expansion of plasma is determined by the equations

$$l_k(t)\frac{d^2 l_k}{dt^2} = \frac{1}{3}\sum_{\alpha} \mu_\alpha V_{k\alpha}^2(t), \tag{55}$$

whose analysis shows that the evolution of the scales $l_k(t)$ strongly depends on the initial ratio of these scales. In particular, if at $t=0$ one of the characteristic dimensions of the plasma bunch is small compared to the other dimensions [e.g., $l_x(0) \ll l_y(0), l_z(0)$], with the passage of time the thermal energy of the plasma is transformed primarily into the energy of collective (hydrodynamic) motion of the particles in this specific direction (x). Accordingly, the expansion of the plasma proceeds in the same direction:

$$\frac{dl_x}{dt} \gg \frac{dl_y}{dt}, \frac{dl_z}{dt}; \quad l_x(t) \gg l_y(t), l_z(t) \ \text{ as } \ t \to \infty.$$

The self-similar solutions of the kinetic equations (1) in this case of a dense plasma can be written[15,16]

$$f_\alpha(\mathbf{r},\mathbf{v},t) = \lambda_\alpha F\left(\sum_{k=1}^{3}\left[\frac{\tilde{r}_k^2}{l_k^2(t)} + \frac{3(\tilde{v}_k - \tilde{u}_k)^2}{V_\alpha^2(t)}\right]\right),$$

$$\tilde{u}_k(\tilde{r}_k, t) = \frac{\tilde{r}_k}{l_k}\frac{dl_k}{dt}, \quad V_\alpha(t)\left[\prod_{k=1}^{3} l_k(t)\right]^{2/3} = \text{const}, \tag{56}$$

where F is an arbitrary function, and the λ_α are normalization constants.

If collisions ensuring that condition (51) is met do not bring about effective exchange of thermal energy between the particles of the different species, the functions $V_\alpha(t)$ can be assumed to be independent for each component of the plasma. Such a situation is realized, e.g., in the expansion of electron–ion plasma with cold heavy ions: $\mu_e \ll \mu_i$ and $\mu_i V_i^2(0) \ll \mu_e V_e^2(0)$ (see Refs. 15 and 16). In the opposite limiting case, where as a result of collisions thermal equilibrium between particles of the different species sets in, we must assume that

$$\mu_\alpha V_\alpha^2(t) = \frac{1}{2}\sum_{\beta=1,2} \mu_\beta V_\beta^2(t).$$

Here the velocity distribution of the particles corresponds to the Maxwellian distribution, i.e., $F(G) \propto \exp(-G)$.

5. DYNAMICS OF A COLLISIONLESS TWO-COMPONENT PLASMA BUNCH IN A SLOWLY VARYING MAGNETIC FIELD

The results of the analysis of the dynamics of collisionless two-component plasma in external slowly varying potential fields U_α may serve as a basis for studying the problem of acceleration and compression of a plasma bunch in a magnetic field with a magnetic-mirror configuration:

$$\mathbf{B}(\mathbf{r},t) = B_0(z,t)\mathbf{e}_z + B_\rho(\rho,z,t)\frac{\mathbf{e}_x + \mathbf{e}_y}{\sqrt{2}}. \tag{57}$$

Here $\rho^2 \equiv x^2 + y^2$, and $\mathbf{B}(\mathbf{r},t)$ is the induction vector of the axisymmetric magnetic field directed mainly along the symmetry axis z:

$$|B_0(z,t)| \gg |B_\rho(\rho,z,t)|, \quad B_\rho(\rho,z,t) = -\frac{\rho}{2}\frac{\partial B_0}{\partial z}.$$

We assume that the magnetic field is large in magnitude and slowly varies in space and time:

$$\omega_\alpha \tau \gg 1, \quad \left|\frac{\dot{\omega}_\alpha}{\omega_\alpha^2}\right| \ll 1, \quad \left|\frac{\partial \omega_\alpha}{\partial z}\right| \ll \left|\frac{\omega_\alpha}{l_z}\right|,$$

where

$$\omega_\alpha \equiv \frac{Z_\alpha e}{m_\alpha c} B_0(z,t)$$

is the cyclotronfrequency of particles of species α in the magnetic field, c is the speed of light, τ is the characteristic expansion time of a plasma bunch, and l_z is the longitudinal plasma size. Then the transverse motion (in relation to the direction of the magnetic field) of the particles can be described in the adiabatic approximation, according to which the distribution function for particles of the α species averaged over the period of cyclotron rotation is

$$f_\alpha = \Phi_\alpha(t, z, v_z, J_\alpha, G_\alpha),$$

$$G_\alpha \equiv \omega_\alpha(x^2 + y^2) + 2(x v_y - y v_x), \quad J_\alpha = \frac{v_x^2 + v_y^2}{\omega_\alpha}, \tag{58}$$

where Φ_α satisfies the kinetic equation

$$\frac{\partial \Phi_\alpha}{\partial t} + v_z\frac{\partial \Phi_\alpha}{\partial z} - \frac{Z_\alpha e}{m_\alpha}\frac{\partial \varphi}{\partial z}\frac{\partial \Phi_\alpha}{\partial v_z} - \frac{1}{2}\frac{\partial \omega_\alpha}{\partial z} J_\alpha \frac{\partial \Phi_\alpha}{\partial v_z} = 0. \tag{59}$$

In the particular case $\Phi_\alpha(t=0) = \Phi_{\alpha 0}(z, v_z)\delta(J_\alpha - J_{\alpha 0})$, $J_{\alpha 0} = \text{const}$, by integrating Eq. (59) over the transverse coordinates and velocities the problem can be reduced to the one-dimensional analog of the problem of the dynamics of a plasma bunch in slowly varying external fields:

$$\frac{\partial F_\alpha}{\partial t} + v_z\frac{\partial F_\alpha}{\partial z} - \frac{Z_\alpha e}{m_\alpha}\frac{\partial \varphi}{\partial z}\frac{\partial F_\alpha}{\partial v_z} - \frac{1}{m_\alpha}\frac{\partial U_\alpha}{\partial z}\frac{\partial F_\alpha}{\partial v_z} = 0,$$

$$N_\alpha F_\alpha(t,z,v_z) \equiv \int\int \Phi_\alpha \, dx\, dy\, dv_x\, dv_y, \quad U_\alpha \equiv g_\alpha B_0(z,t), \tag{60}$$

$$g_\alpha \equiv \frac{m_\alpha}{2B_0}\frac{\int\int (v_x^2 + v_y^2)\Phi_\alpha \, dx\, dy\, dv_x\, dv_y}{\int\int \Phi_\alpha \, dx\, dy\, dv_x\, dv_y},$$

where the quantity g_α, defined at $t=0$, can be set constant in the adiabatic approximation.

Thus, in accordance with the results of Sec. 3, the motion of the center of mass of the plasma bunch ($Z \equiv \int\int z F_\alpha dz dv_z$) and the dynamics of the characteristic longitudinal plasma size ($l_z^2(t) \equiv \int\int (z - Z(t))^2 F_\alpha \, dz\, dv_z$) in a slowly varying magnetic field are determined by the equations

$$\frac{d^2 Z}{dt^2} = -\gamma \left.\frac{\partial B_0}{\partial z}\right|_{z=Z(t)}, \quad \gamma \equiv \frac{\Sigma_\alpha N_\alpha g_\alpha}{M \Sigma_\alpha m_\alpha}, \tag{61}$$

$$l_z^3\frac{d^2l_z}{dt^2}+\Omega^2(t)l_z^4=V_z^2l_z^2=\text{const},\quad \Omega^2(t)\equiv\gamma\frac{\partial^2B_0}{\partial z^2}\bigg|_{z=Z(t)}, \tag{62}$$

where $V_z^2\equiv 2W_z/M-(dl_z/dt)^2$, and W_z is the kinetic energy of the particle motion in the longitudinal direction in the center-of-mass reference frame.

Equations (61) and (62) make it possible to study the acceleration of a plasma bunch in a moving magnetic mirror. According to Eq. (61), the plasma bunch is pushed out of the region where the magnetic field is strong, i.e., for the bunch to be accelerated in the positive direction of the z axis, we must ensure that $\partial B_0/\partial z<0$. At the same time, Eq. (62) implies that for the longitudinal plasma size to be conserved the second derivative of B_0 must be positive:

$$\frac{\partial^2B_0}{\partial z^2}=\frac{V_z^2}{\gamma l_z^2}.$$

Thus the possibilities for the acceleration of a plasma bunch as a whole in a time-independent magnetic field are extremely limited. These possibilities can be broadened by using a traveling magnetic-field wave:

$$B_0=\tilde{B}_0(t)+\frac{\Omega^2}{2\gamma}\left(z-Z_0(t)-\frac{a(t)}{\Omega^2}\right)^2, \tag{63}$$

where $\tilde{B}_0(t)$ and $a(t)=d^2Z_0/dt^2$ are arbitrary functions of time, and Ω^2 is a positive constant. Such a wave guarantees a given acceleration of the bunch $[Z(t)=Z_0(t)]$ with the longitudinal size conserved:

$$l_z^2=\frac{2W_z}{M\Omega^2}=\text{const}. \tag{64}$$

The variation of the characteristic transverse plasma scale in this case is determined by the adiabatic law $l_\rho^2B_0(Z_0,t)=\text{const}$, i.e.,

$$l_\rho^2\propto\frac{1}{\tilde{B}_0(t)+\dfrac{a^2(t)}{2\gamma\Omega^2}}.$$

Note that this acceleration regime is stable. In other words, if at $t=0$ the coordinate Z and the velocity dZ/dt of the center of mass of the plasma bunch differ from the initial values Z_0 and dZ_0/dt,

$$Z(t)-Z_0(t)=\xi(t),\quad \xi(0)\neq 0,\quad \frac{d\xi}{dt}\neq 0.$$

the function $\xi(t)$ does not increase with time since it satisfies the harmonic-oscillator equation

$$\frac{d^2\xi}{dt^2}+\Omega^2\xi=0. \tag{65}$$

The longitudinal plasma size $l_z(t)$ experiences similar oscillations if condition (64) is not met initially:

$$\frac{d^2\eta}{dt^2}+4\Omega^2\eta=0,\quad \eta\equiv l_z^2-\frac{2W_z}{M\Omega^2}. \tag{66}$$

Using Eqs. (61) and (62), we can study the compression of a plasma bunch in an exploding magnetic trap. If the magnetic field varies in time according to the law

$$B_0(z,t)=\tilde{B}_0(t)+\frac{\Omega^2(t)}{2\gamma}z^2,\quad \Omega^2(t)>0, \tag{67}$$

where $\tilde{B}_0(t)$ and $\Omega^2(t)$ are slowly increasing functions of time, the compression of the bunch is described by the adiabatic laws

$$l_\rho^2\tilde{B}_0(t)=\text{const},\quad \frac{\overline{W}_\rho}{\tilde{B}_0(t)}=\text{const}, \tag{68}$$

$$l_z^2\Omega(t)=\text{const},\quad \frac{\overline{W}_z}{\Omega(t)}=\text{const}, \tag{69}$$

where $\overline{W}_\rho$ and $\overline{W}_z$ are the energies of thermal motion of the particles in the longitudinal and transverse directions, respectively.

6. CONCLUSION

Our results show that the method of moments is an effective instrument in theoretical studies of the dynamics of quasineutral plasma bunches in external fields. Its use makes it possible to calculate, in the general case, the evolution of the characteristic dimensions of the bunch and the cooling of a multicomponent plasma in the process of plasma expansion into vacuum. In the simplest case of a two-component collisionless plasma, the method of moments provides a complete description of the dynamics of the bunch in slowly varying external potential fields, which can be used as a basis for finding analytical solutions of the appropriate kinetic equations. These solutions can be found for an arbitrary mass ratio of the constituent particles and for arbitrary velocity distributions of these particles in a broad class of spatial plasma-density distributions. We believe that the simplicity and efficiency of the method argue in favor of its use in solving many applied problems.

This work was made possible by Grants from the Russian Fund for Fundamental Research (Grant No. 98-02-17052) and the Controlled Fusion Program (Grant No. 369).

*)E-mail: dorozh@appl.sci-nnov.ru
†)E-mail: sss@appl.sci-nnov.ru

1)In Eqs. (21)–(23) we have discarded the subscript α on the spatial and mixed moments, since they coincide for particles of the two species.
2)In the absence of external fields ($U_\alpha=0$), Eq. (36) becomes Eq. (24).
3)Examples of such processes are various types of plasma instability.

[1] A. V. Gurevich, L. V. Pariĭskaya, and L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **49**, 644 (1965) [Sov. Phys. JETP **22**, 449 (1966)].
[2] A. Gurevich, D. Anderson, and H. Wilhelmsson, Phys. Rev. Lett. **42**, 769 (1979).
[3] A. V. Gurevich and A. P. Meshcherkin, Zh. Éksp. Teor. Fiz. **80**, 1810 (1981) [Sov. Phys. JETP **53**, 937 (1981)].
[4] A. V. Gurevich and A. P. Meshcherkin, Zh. Éksp. Teor. Fiz. **81**, 1295 (1981) [Sov. Phys. JETP **54**, 688 (1981)].
[5] Y. El-Zein, A. Amin, H. S. Kim *et al.*, Phys. Plasmas **2**, 1073 (1995).

[6] S. Sakabe, T. Mochizuki, T. Yabe, K. Mima, and C. Yamanaka, Phys. Rev. A **26**, 2159 (1982).
[7] A. V. Gurevich and A. P. Meshcherkin, Fiz. Plazmy **9**, 955 (1983) [Sov. J. Plasma Phys. **9**, 556 (1983)].
[8] Ch. Sack and H. Schamel, Phys. Rep. (Rev. Sect. Phys. Lett.) **156**, 311 (1987).
[9] G. Manfredi, S. Mola, and M. R. Feix, Phys. Fluids B **5**, 388 (1993).
[10] L. G. Garcia, J. Goedert, H. Figua, E. Fijalkow, and M. R. Feix, Phys. Plasmas **4**, 4240 (1997).
[11] A. V. Baitin and K. M. Kuzanyan, J. Plasma Phys. **59**, 83 (1998).
[12] D. S. Dorozhkina and V. E. Semenov, Fiz. Plazmy **24**, 481 (1998) [Plasma Phys. Rep. **24**, 440 (1998)].
[13] D. S. Dorozhkina and V. E. Semenov, JETP Lett. **67**, 573 (1998).
[14] D. S. Dorozhkina and V. E. Semenov, Phys. Rev. Lett. **81**, 2691 (1998).
[15] D. S. Dorozhkina and V. E. Semenov, in *Proc. Conf. on Low-Temperature Plasma Physics*, *FNTP-98*, Petrozavodsk (1998), Part 1, p. 486.
[16] D. S. Dorozhkina and V. E. Semenov, in *Proc. 1998 Int. Conf. on Plasma Physics combined with 25 EPS Conference on Controlled Fusion and Plasma Physics*, Prague (1998), Vol. 22C, p. 285.

Translated by Eugene Yankovsky

Dust acoustic waves in a dc glow-discharge plasma

V. I. Molotkov,*) A. P. Nefedov, V. M. Torchinskiĭ, V. E. Fortov, and A. G. Khrapak

Scientific-Research Center for the Thermal Physics of Pulsed Effects, Russian Academy of Sciences, 127412 Moscow, Russia

(Submitted 10 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 902–907 (September 1999)

The spontaneous excitation of low-frequency oscillations of the macroparticle density in ordered dust structures levitating in standing striations of a dc glow discharge is discovered. It is concluded on the basis of a simplified linear model of an ideal collisionless plasma that the observed instability is caused by the drift motion of ions relative to the dust, which leads to the excitation of dust acoustic oscillations of the plasma. © *1999 American Institute of Physics.* [S1063-7761(99)01109-9]

The presence of charged dust particles in a low-temperature plasma leads to the appearance of new oscillation modes and instabilities.[1–3] For example, the phase velocity of ion sound increases as a result of the decrease in the concentration of electrons, which are partially absorbed by the macroparticles. This leads to changes in the characteristics of the ion-acoustic current instability appearing because of the relative motion of ions and electrons at frequencies close to the ion plasma frequency. At lower frequencies, close to the dust plasma frequency, current instability can appear as a result of the motion of electrons and ions relative to the charged dust particles (see Ref. 3 and the literature cited therein). Dust sound and a corresponding current instability were recently observed in laboratory experiments.[4–8] Dust acoustic instability can appear in various systems, such as, for example, Saturn's rings, radio-frequency discharges used in plasma-sputtering and etching technologies, and plasma crystals.[3]

The appearance of natural oscillations in a dusty dc glow-discharge plasma sustained in neon was discovered in the present work, and an attempt was made to interpret this phenomenon as being a result of a plasma-dust current instability. The experimental setup scarcely differed from the one which we previously used in Ref. 9. The plasma-dust structures were formed in standing striations of a low-pressure discharge in a glass tube with a diameter of 3 cm and cold electrodes. Monodisperse microspheres of a melamine-formaldehyde resin ($\rho = 1.5\ \mathrm{g/cm^3}$) with diameters of 10.24 and 1.87 μm, whose charge ranged from 10^5 to $10^4\,e$, were used in the experiments. The structures were visualized using transillumination by a laser ''knife'' in a vertical plane. Video images of the structures were recorded using a CCD camera and a video cassette recorder. Figure 1 presents a video image of a structure consisting of particles with a diameter of 1.87 μm. Oscillations of the dust particle density are clearly seen in the lower part of the structure in the video image. These oscillations are particle density waves with a wavelength $L \sim 1$ mm and an oscillation period $T \sim 5 \times 10^{-2}$ s, which travel downward from the anode to the cathode. It should be stressed that these oscillations exist only in the lower part of the structures, whose linear dimension and position correspond to the head of a striation, where the electric field intensity is greatest. In addition, it was discovered that the oscillations appear when there is a definite (critical) number of dust particles in the structure. This can be seen in Fig. 2: the first frame [Fig. 2(a)] shows a well ordered structure, and the ensuing frames [Figs. 2(b) and 2(c)] show the development of instability in response to the additional injection of particles and their trapping by the structure.

We note that the oscillations disappear when the discharge current is raised or the gas pressure is increased.

The frequency of the oscillations discovered is close to the frequency of plasma-dust oscillations.[3] Therefore, an explanation for the effects described above should be sought in the possible instabilities of the low-frequency oscillations of a dusty plasma. The spectrum of longitudinal modes of a plasma is determined from the solution of the dispersion equation

$$\varepsilon(\omega, \mathbf{k}) = 0, \tag{1}$$

where ε is the dielectric constant of the plasma, and ω and $\mathbf{k}$ are the frequency and wave vector of the oscillations. The susceptibility of an ideal motionless plasma $\chi = \varepsilon - 1$ is additive with respect to the charged components of the plasma:

$$\varepsilon(\omega, \mathbf{k}) = 1 + \sum_{j=e,i,d} [\varepsilon^j(\omega, \mathbf{k}) - 1]. \tag{2}$$

Here the indices e, i, and d correspond to electrons, ions, and dust particles. In a gas discharge the velocity distribution of the charged particles deviates from equilibrium because of the directed motion in the electric field with the drift velocities $\mathbf{u}_j$. The dielectric constant ε^j of each of the components in a coordinate frame moving with the velocity $\mathbf{u}_j$ has the same form as in the laboratory coordinate frame with $\mathbf{u}_j = 0$. In going over to the laboratory frame, allowance should be made for the Doppler frequency shift, which leads to generalization of the expression (2) to the case of nonzero drift velocities:

$$\varepsilon(\omega, \mathbf{k}) = 1 + \sum_{j=e,i,d} [\varepsilon^j(\omega - \mathbf{k}\cdot\mathbf{u}_j, \mathbf{k}) - 1]. \tag{3}$$

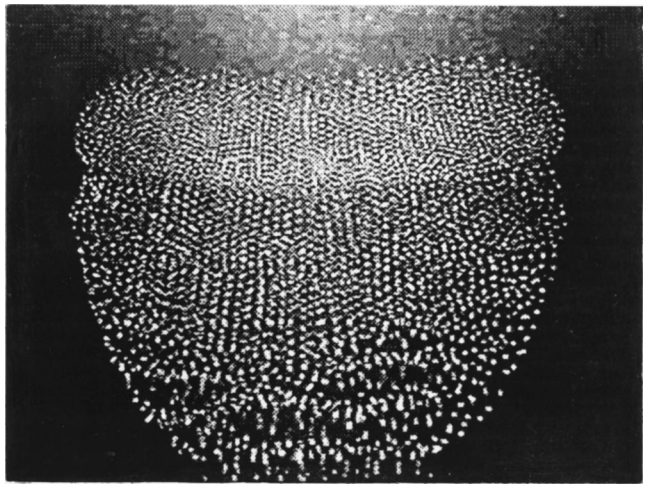

FIG. 1. Video image of an ordered structure of monodisperse particles with a diameter of 1.87 μm at a discharge current of 5 mA and a pressure of 0.3 Torr. Each frame corresponds to 10.6 mm in the vertical direction.

In the case of a collisionless Maxwellian plasma in the absence of a magnetic field, the solution of Vlasov's equation leads to the following expression for the longitudinal dielectric constant:[10]

$$\varepsilon^j(\omega,\mathbf{k})=1+\frac{1}{(k\lambda_j)^2}\left[1+F\left(\frac{\omega}{\sqrt{2}kv_j}\right)\right], \tag{4}$$

where the parameters

$$\lambda_j=\sqrt{\frac{T_j}{4\pi N_j e^2}}, \quad v_j=\sqrt{\frac{T_j}{m_j}} \tag{5}$$

are the Debye length and the mean thermal velocity of the jth component, and T_j, N_j, and m_j are the temperature, concentration, and mass of the particles of the jth component. The function $F(x)$ is defined by the integral

$$F(x)=\frac{x}{\sqrt{\pi}}\int_{-\infty}^{\infty}\frac{\exp(-z^2)dz}{z-x-i0}$$
$$\simeq\begin{cases}-1-\dfrac{1}{2x^2}-\dfrac{3}{4x^4}+i\sqrt{\pi}\,xe^{-x^2}, & x\gg 1,\\ -2x^2+i\sqrt{\pi}\,x, & x\ll 1.\end{cases} \tag{6}$$

In laboratory experiments dust particles levitate and perform chaotic thermal motions, and their drift velocity u_d is equal to zero. The thermal velocity of the electrons v_e is usually significantly greater than their drift velocity u_e, and the latter can also be considered equal to zero. The following inequalities usually hold in a dc gas-discharge plasma in the region of parameters where dust acoustic instabilities are observed:

$$kv_e\gg kv_i>ku_i\gg\omega\gg kv_d. \tag{7}$$

Thus, in accordance with (3)–(6), the complex dielectric constant can be represented in the form

$$\varepsilon(\omega,\mathbf{k})=1-\frac{\omega_d^2}{\omega^2}+\frac{1}{k^2\lambda^2}+i\sqrt{\frac{\pi}{2}}\frac{\omega-u_ik}{k^3v_i\lambda_i^2}, \tag{8}$$

where

$$\omega_d=\left(\frac{4\pi N_dZ_d^2e^2}{m_d}\right)^{1/2}, \quad \lambda=\frac{\lambda_e\lambda_i}{\sqrt{\lambda_e^2+\lambda_i^2}} \tag{9}$$

are the dust plasma frequency and the electron-ion Debye length, respectively, and Z_d is the charge of the dust particles.

Assuming that the absolute value of the imaginary part of ε is small and setting

$$\omega=\omega_r+i\gamma, \quad \omega_r\gg\gamma, \tag{10}$$

from (1) and (8) we find the low-frequency oscillation spectrum of a dusty plasma:

$$\omega_r^2\simeq\omega_d^2\frac{k^2\lambda^2}{1+k^2\lambda^2}, \tag{11}$$

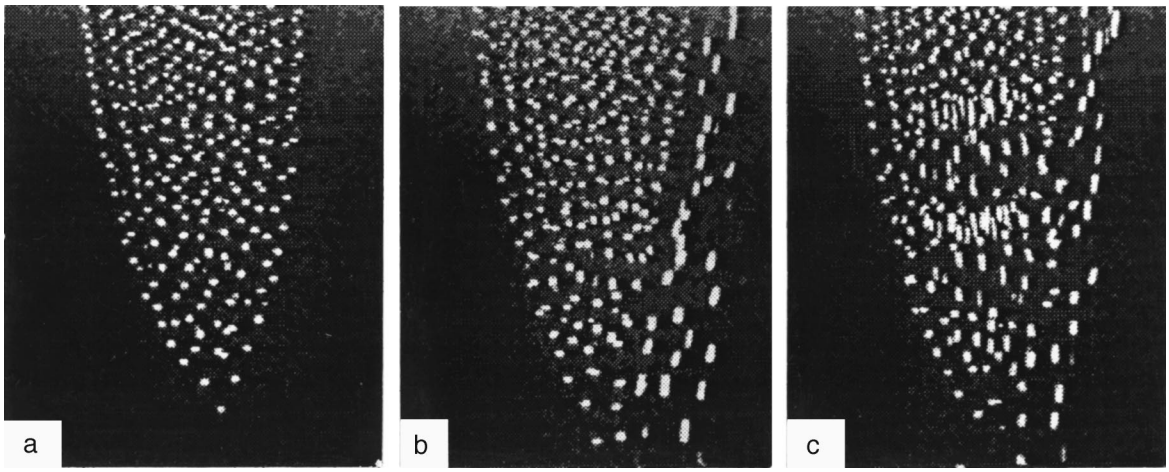

FIG. 2. Video image of fragments of structures of monodisperse particles with a diameter of 1.87 μm at a discharge current of 0.6 mA and a pressure of 0.3 Torr. Each frame corresponds to 6 mm in the vertical direction.

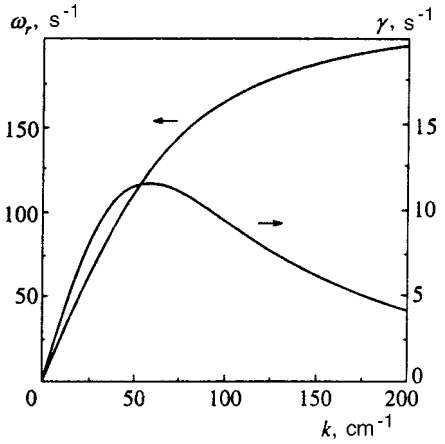

FIG. 3. Dispersion $\omega_r(k)$ and growth rate $\gamma(k)$ of low-frequency plasma-dust oscillations in standing striations of a dc gas discharge.

$$\gamma \simeq -\sqrt{\frac{\pi}{8}}\,\frac{\omega_r^3}{\omega_d^2 k^3 \lambda_i^2}\,\frac{\omega_r - u_i k}{\nu_i}. \tag{12}$$

When $u_i = 0$, this spectrum coincides with the spectrum of dust acoustic oscillations. A nonzero value of the drift velocity of the ions u_i leads to a decrease in the damping decrement γ, and at values of u_i exceeding the phase velocity of the waves $v_{\rm ph} = \omega/k$, the damping decrement γ changes sign, i.e., instability appears. In complete analogy to the ion acoustic instability of an ordinary plasma,[11] the instability discovered is caused by the Cherenkov radiation of dust acoustic waves by ions moving with a supersonic velocity. It is possible only under the conditions

$$Z_d T_i \gg T_d, \quad u_i > v_{\rm ph} \gg v_d, \tag{13}$$

which are satisfied with a large safety margin in a dusty dc glow-discharge plasma.

Under the conditions of our experiment, at a pressure $p \simeq 1$ Torr and an electric field intensity $E \simeq 3$ V/cm, the ion drift velocity u_i is roughly equal to 8×10^3 cm/s. For a characteristic oscillation frequency $\omega = 2\pi/T \simeq 60\ {\rm s}^{-1}$ and a wave vector $k = 2\pi/L \simeq 60\ {\rm cm}^{-1}$ the phase velocity of the waves is small compared with the ion drift velocity: $v_{\rm ph} = \omega/k \simeq 1$ cm/s $\ll u_i$. Estimates made in accordance with (9) and (5) for particles with a diameter of 1.87 μm give $\omega_d \simeq 210\ {\rm s}^{-1}$ and $\lambda \simeq 1.2 \times 10^{-2}$ cm (the values $Z_d \simeq 2.5 \times 10^3\, e$, $N_d \simeq 10^4\ {\rm cm}^{-3}$, and $N_i \simeq 10^8\ {\rm cm}^{-3}$ were used). The results of the calculation of the frequency ω_r and the growth rate γ of the dust acoustic oscillations are presented in Fig. 3. The instability growth rate has a maximum at $k = k_m = 1/\sqrt{2}\lambda \simeq 60\ {\rm cm}^{-1}$ at the characteristic frequency $\omega_r(k_m) = \omega_d/\sqrt{3} \simeq 120\ {\rm s}^{-1}$. Just such waves are excited in our experiment.

Despite the good agreement with experiment, the model proposed above cannot claim to provide a faithful quantitative description of the spectrum of dust acoustic oscillations, since the linear theory of an ideal collisionless plasma was used to substantiate it. Nevertheless, it provides explanations for several qualitative features of the phenomenon observed. For example, the development of the instability only in the lower part of the dust structure resting on the head of a striation is probably due to the fact that the electric field E and thus the ion drift velocity u_i reach maxima in this region. According to (12), the instability growth rate also reaches its greatest value in this region.

In our opinion, dust acoustic instability is not observed in radio-frequency discharges, because in the layers near the electrodes of these discharges, where the levitation of dust particles is usually observed, due to the Bohm effect[12] the ion drift velocity satisfies the condition

$$u_i \simeq \sqrt{\frac{T_e}{m_i}} \gg \nu_i = \sqrt{\frac{T_i}{m_i}}.$$

This leads to alteration of the spectrum of dust acoustic oscillations (11) and the appearance of an exponentially small multiplier $\exp(-T_e/T_i)$ in the instability growth rate (12). The recent discovery of dust acoustic oscillations in a radio-frequency discharge under microgravitational conditions[13] does not contradict the foregoing statements, since in this case dust structures are located throughout the volume of the plasma and the phenomenon under consideration occurs far from the electrodes, where $u_i < v_i$.

Dust acoustic instability can be initiated by the decrease in the gas pressure in the discharge or by the increase in the number of macroparticles in the dust structure. The former effect is associated with an increase in the ion drift velocity and a decrease in the viscosity of the neutral gas. The latter effect, which is illustrated in Fig. 2, possibly occurs because the increase in the concentration of dust particles creates an additional channel for a loss of charges (apart from the principal channel associated with ambipolar diffusion on the walls of the discharge tube), which, at a fixed discharge current, necessitates an increase in the ionization frequency and, consequently, leads to intensification of the field in the region where the dust particles are found.[9] This, in turn, leads to a rise in the ion drift velocity u_i and, as a result, to an increase in the instability growth rate.

Finally, we note that the disappearance of the oscillations in response to an increase in the discharge current is probably a consequence of the lowering of the electric field intensity ordinarily observed under such conditions.

We thank S.A. Khrapak for some useful discussions.

*)E-mail: molotkov@hedric.msk.su

[1] M. Rosenberg, in *The Physics of Dusty Plasmas*, P. K. Shukla, D. A. Mendis, and V. W. Chow (Eds.), World Scientific, Singapore (1996), p. 129.
[2] F. Verheest, Space Sci. Rev. **77**, 267 (1996).
[3] D. Winske and M. Rosenberg, IEEE Trans. Plasma Sci. **26**, 92 (1998).
[4] A. Barkan, R. L. Merlino, and N. D'Angelo, Phys. Plasmas **2**, 3563 (1995).
[5] G. Praburam and J. Goree, Phys. Plasmas **3**, 1212 (1996).
[6] C. Thompson, A. Barkan, N. D'Angelo, and R. L. Merlino, Phys. Plasmas **4**, 2331 (1997).
[7] V. V. Zhakhovskiĭ, V. I. Molotkov, A. P. Nefedov, V. M. Torchinskiĭ, V. E. Fortov, A. G. Khrapak, in *Reports of the Conference on Low-Temperature Plasma Physics (LTPP-98)* [in Russian], Izd. Petrozavodsk. Univ., Petrozavodsk (1998), p. 684.

[8] R. L. Merlino, A. Barkan, C. Thompson, and N. D'Angelo, Phys. Plasmas **5**, 1607 (1998).
[9] A. M. Lipaev, V. I. Molotkov, A. P. Nefedov, O. F. Petrov, V. M. Torchinskiĭ, V. E. Fortov, A. G. Khrapak, and S. A. Khrapak, Zh. Éksp. Teor. Fiz. **112**, 2030 (1997) [JETP **85**, 1110 (1997)].
[10] E. M. Lifshitz, L. P. Pitaevskiĭ, and L. D. Landau, *Physical Kinetics*, Pergamon Press, Oxford (1981) [Russ. original, Nauka, Moscow (1979)].
[11] V. L. Ginzburg and A. A. Rukhadze, *Waves in a Magnetically Active Plasma* [in Russian], Nauka, Moscow (1970).
[12] M. A. Liberman and A. J. Lichtenberg, *Principles of Plasma Discharges and Materials Processing*, Wiley, New York (1994).
[13] G. E. Morfill and H. Thomas (1998) (private communication).

Translated by P. Shelnitz

Three-dimensional array structures associated with Richtmyer–Meshkov and Rayleigh–Taylor instability

N. A. Inogamov*)

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

A. M. Oparin†)

Institute of Computer-Aided Design, Russian Academy of Sciences, 123056 Moscow, Russia

(Submitted 20 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 908–939 (September 1999)

A boundary separating adjacent gas or liquid media is frequently unstable. Richtmyer–Meshkov and Rayleigh–Taylor instability cause the growth of intricate structures on such boundaries. All the lattice symmetries [rectangular ($pmm2$), square ($p4mm$), hexagonal ($p6mm$), and triangular ($p3m1$) lattices] which are of interest in connection with the instability of the surface of a fluid are studied for the first time. They are obtained from initial disturbances consisting of one (planar case, two-dimensional flow), two (rectangular cells), or three (hexagons and triangles) harmonic waves. It is shown that the dynamic system undergoes a transition during development from an initial, weakly disturbed state to a limiting or asymptotic stationary state (stationary point). The stability of these points (stationary states) is investigated. It is shown that the stationary states are stable toward large-scale disturbances both in the case of Richtmyer–Meshkov instability and in the case of Rayleigh–Taylor instability. It is discovered that the symmetry increases as the system evolves in certain cases. In one example the initial Richtmyer–Meshkov or Rayleigh–Taylor disturbance is a sum of two waves perpendicular to one another with equal wave numbers, but unequal amplitudes: $a_1(t=0)\neq a_2(t=0)$. Then, during evolution, the flow has $p2$ symmetry (rotation relative to the vertical axis by 180°), which goes over to $p4$ symmetry (rotation by 90°) at $t\to\infty$, since the amplitudes equalize in the stationary state: $a_1(t=\infty)=a_2(t=\infty)$. It is shown that the hexagonal and triangular arrays are complementary. Upon time inversion ($t\to -t$), ''rephasing'' occurs, and the bubbles of a hexagonal array transform into jets of a triangular array and vice versa.
 [S1063-7761(99)01209-3]

1. INTRODUCTION

Richtmyer–Meshkov instability appears when a shock wave passes through a boundary between media of different density, and Rayleigh–Taylor instability is generated by unstable gravitational stratification. The stirring of the media caused by them is significant in many physical problems, for example, in laser and beam inertial confinement fusion,[1–3] astrophysics,[4,5] and the physics of explosions.[6] The corresponding questions have been studied intently, and it is sufficient to cite the latest studies.[3,7–10] The theory of Richtmyer–Meshkov and Rayleigh–Taylor instability is quite formidable, being comparable in complexity to the theory of gravity waves.

Review of the theory. Let us recall the principal theoretical studies,[1)] grouping them according to the methods used. The two-dimensional (2D) case has been investigated for the most part, since the three-dimensional (3D) case is far more complicated. Therefore, we single out the studies in which the 3D case was investigated.

1. Parabolic model, Richtmyer–Meshkov and Rayleigh–Taylor instability, nonstationary and stationary cases. The Layzer model,[11] which is based on the parabolic approximation of the boundary ($\eta=\cdots x^{2N}$, $N=1$) near a bubble apex, is effective. It leads to a dynamic system, whose trajectories describe the gradual transformation of the boundary strains from the original, weakly disturbed state to the establishment of a stationary state (or stationary point). The stationary state is achieved asymptotically as $t\to\infty$. The scope of the problem can be narrowed, and the stationary state can be studied at once (the derivatives with respect to time are set equal to zero at the stationary point). Then, instead of a dynamic system and differential equations, there is a system of algebraic equations, which represents a stationary version of the Layzer model. A first-order stationary state (parabolic approximation, $N=1$) was studied back before Layzer's work in the notable work by Davies and Taylor.[12] The Layzer approach is applicable to both Richtmyer–Meshkov instability ($g=0$, g is the acceleration of free fall) and Rayleigh–Taylor instability ($g=1$).[13–15] If the case of a rectangular lattice is ruled out (for further information on this lattice and other 3D generalizations, see below), the phase space of the Layzer model is planar.[15] This permits the investigation of the corresponding dynamic system in the most general case[15] (see also the subsequent publications in Refs. 7, 9, and 16–18). The analysis demonstrates, in particular, the stability of

Richtmyer–Meshkov and Rayleigh–Taylor stationary states in the class of large-scale disturbances, since each stationary point is a node.[15,17,18]

2. Functional vicinity of the equilibrium configuration. Several studies have been carried out in the weakly nonlinear approximation, which is valid for small disturbance amplitudes (see Refs. 19 and 20 and the references therein). This approximation describes only the beginning of the departure from the bubble-jet stage of motion, which is symmetric with respect to inversion,[2)] toward bubble-jet asymmetry. This asymmetry is manifested by broadening of the bubble and narrowing of the jet.

3. Integral formulation of the boundary-value problem of potential theory. The results determined can be obtained using conformal transformations after representing the boundary-value problem in the form of an integral equation.[21–23]

4. Successive approximations. Higher generalizations of model 1 in the Nth-order approximation ($\eta=\cdots x^{2N}$, where N is fairly large) were developed in Refs. 13, 15, 17, and 24. They refer to both the stationary[17,24] and nonstationary[13,15] cases. Higher expansions provide a powerful tool for investigating convergence and uniqueness problems.[17]

Combining approaches 1 and 4. An obvious advantage of method 1 over 2 and 3 is that in 1 the transformation from the linear limit to the nonlinear limit or an asymptotic state can be traced in its entirety. When it is combined with approach 4, possibilities are opened up for obtaining exhaustive results on Richtmyer–Meshkov and Rayleigh–Taylor instability. A significant shortcoming is the poor extent of development of the higher stationary approximations for Richtmyer–Meshkov instability (the stationary boundary conditions are complicated, and it is not clear whether the states sought are zero- or one-parameter).

Combined approach in the 3D case. The extension of method 1 to the 3D geometry (see below) gives a general description of the evolution. It turns out that it is accompanied by striking structural effects.[3)] For its part, the extension of technique 4, which was developed mainly for the 2D case, provides important qualitative conclusions regarding the convergence and uniqueness of Rayleigh–Taylor instability in the 3D case. Thus, in this paper we are dealing with the extension of methods 1 and 4 to the 3D geometry.

Previous studies on 3D flows. In Refs. 17, 28, and 29 method 4 was applied to the stationary stage of Rayleigh–Taylor instability in the 3D case. The values of N were small. References 30–32 were devoted to three-dimensional numerical simulation. Some 3D experiments in shock tubes were described in Refs. 33 and 34, and similar experiments in an explosive system were described in Ref. 35.

Content. Sections 2 and 3 present a statement of the problem, the boundary conditions, and the spectral decomposition of the 3D potentials. The symmetry of the potentials and the unit cells of the array is discussed. Sections 4 and 9 describe the derivation of the equations for generalizing model 1 to the 3D case. A complete analysis of the corresponding phase space is given. Exact integrals of the expressions obtained are given for the case of Richtmyer–Meshkov instability. Sections 4, 5, and 7 explore the 3D structure of the surface η. Section 6 presents the results of a direct numerical simulation. The Rayleigh–Taylor stationary states are presented in Secs. 10 and 11. The one-parameter character of the 3D stationary states is demonstrated (Sec. 10), and it is shown that the stationary point is unique (Sec. 11).

2. HARMONICITY AND BOUNDARY CONDITIONS

Flow is described by the velocity potential φ ($\Delta\varphi=0$, $\mathbf{V}=\nabla\varphi$). The boundary conditions have the form[36]

$$\eta_t=w|-\eta_x u|-\eta_y v|, \quad u=\varphi_x, \quad v=\varphi_y, \quad w=\varphi_z, \tag{2.1}$$

$$-2\varphi_t|=2g\,\eta+u^2|+v^2|+w^2|,$$

$$f|\equiv f|_\eta\equiv f[x,y,z=\eta(x,y,t),t]. \tag{2.2}$$

The fluid boundary is assigned by the function $z=\eta(x,y,t)$. The acceleration is $g=0$ for Richtmyer–Meshkov instability and $g=1$ for Rayleigh–Taylor instability. The kinematic condition (2.1) describes the advection of the surface η by the velocity field $\nabla\varphi$. The dynamic equation (2.2) is the Cauchy–Lagrange integral of Euler's equation. The stationary case is obtained when $\partial_t=0$, and the two-dimensional case is obtained when $\partial_y=0$. In the stationary case the condition (2.2) transforms into Bernoulli's integral.

3. RECTANGULAR, HEXAGONAL, AND TRIANGULAR LATTICES

Decomposition symmetry. The flow symmetry is determined by the symmetry of the potential. The spectral decompositions of the potentials have the form

$$\varphi=-\frac{1}{2}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{a_{nm}}{q_{nm}}c_{nx}c_{mqy}e_{nm}, \tag{3.1}$$

$$c_{nx}=\cos nx, \quad c_{mqy}=\cos mqy,$$

$$e_{nm}=\exp(-q_{nm}z), \quad q_{nm}=\sqrt{n^2+m^2q^2},$$

$$\varphi=-\frac{1}{6}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{a_{nm}}{q_{nm}}(c_nc_m^{+}+s_ns_m^{+}+c_nc_m^{-}+s_ns_m^{-}$$

$$+c_n^{+}c_m^{-}-s_n^{+}s_m^{-})e_{nm}, \tag{3.2}$$

$$c_n=\cos nx, \quad c_n^{\pm}=\cos n\xi^{\pm}, \quad s_n=\sin nx,$$

$$s_n^{\pm}=\sin n\xi^{\pm}, \quad \xi^{\pm}=\frac{x\pm\sqrt{3}y}{2},$$

$$e_{nm}=\exp(-q_{nm}z), \quad q_{nm}=\sqrt{n^2-nm+m^2}.$$

The series (3.1) refers to the rectangular case. The unit cell is a pipe with a generatrix along the z axis and a transverse section $2\pi\times 2\pi/q$ lying in the (x,y) plane (see Fig. 1). The ratio between the sides of the rectangle equals q. An analysis of the rectangular lattice allows us to understand how 2D flow transforms into 3D flow. The series (3.1) is a combination of the two vectors (1,0,0) and $(0,q,0)$, and the series (3.2) is a combination of the three vectors (1,0,0) and $(1/2, \pm\sqrt{3}/2,0)$. In the series (3.1) each term is an eigenmode of

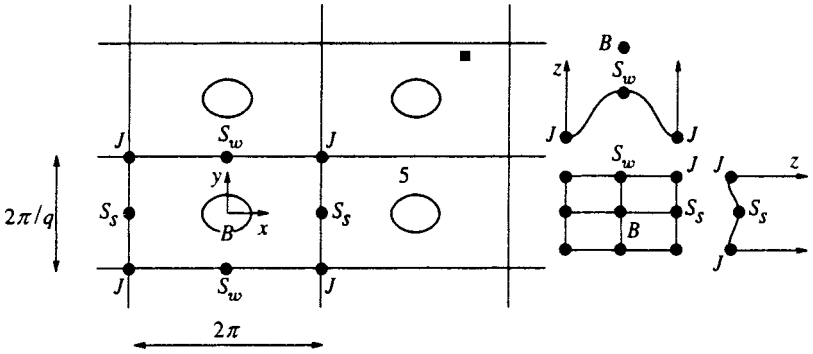

FIG. 1. Rectangular array of bubbles B. Their apices are surrounded by ellipses.

the Laplacian, and in the series (3.2) the pairs $(cc^{\pm}+ss^{\pm})e$ and $(c^{+}c^{-}-s^{+}s^{-})e$ are eigenmodes.

The matrices of the amplitudes a_{nm} (3.1) and (3.2) are functions of time. The amplitudes a_{nm} (3.1) are not symmetric (when $q\neq 1$) with respect to interchange of the indices. The amplitudes a_{nm} (3.2) are symmetric ($a_{nm}=a_{mn}$). A square lattice is obtained when $q=1$ is substituted into the series (3.1). In this case the amplitudes are symmetric.

Unit cell and ideal pipe. The expansions (3.1) and (3.2) of the potentials rapidly converge (this was shown for the 2D case in Refs. 15 and 17). In the series (3.1) a_{10} and a_{01} are dominant,[4)] and in (3.2) a_{10} is dominant. Let us examine the unit cell of the potential (3.1). The center of the transverse coordinates ($x=0, y=0$) is located at point B (Fig. 1). If the initial amplitudes $a_{10}(0)$ and $a_{01}(0)$ are positive and there are no initial disturbances on the surface η, point B is a bubble apex. In Fig. 1 the apices are encircled by ellipses. It is not difficult to see that rectilinear vertical (i.e., parallel to the z axis) trajectories of the fluid particles pass through B (bubble), J (jet), S_s (''strong'' saddle), and S_w (''weak'' saddle) points. Planar trajectory curves of the fluid particles fill the planes which are parallel to the z axis and pass through the straight lines BS_s, BS_w, JS_s, and JS_w. Segments of these straight lines are drawn in the small rectangle in Fig. 1. If $q=1$ (square), planar trajectories also fill the planes passing through the diagonals BJ.

The facts just indicated are associated with the obvious symmetry properties of the potential (3.1). On the symmetric verticals at the B, J, S_s, and S_w points the transverse (horizontal) velocities u and v (2.1) are equal to zero. This is because the BS_s, BS_w, JS_s, and JS_w symmetry planes intersect along these straight lines. The normal component of the velocity V_n vanishes in these planes. Therefore, the quarter of a complete flow period enclosed within rectangle BS_sJS_w is equivalent to flow in an ideal pipe. The normal component of the velocity V_n, rather than the total velocity V, is equal to zero on the walls of such a pipe. The quadrant BS_sJS_w forms a minimal pipe (domain). The flow in the entire space is obtained by parquet multiplication of this domain by reflections and translations. In the case of a square ($q=1$), triangle BJS_s is a domain.

Domains, wall jets, and arches. Near-wall or wall jets travel along the domain walls. The form of these jets is shown schematically in Fig. 1. Curves JS_sJ and JS_wJ correspond to the lower (along z) edges of near-wall jets. They have the form of arches or horseshoes with a dome or a peak at the S_s and S_w saddle points, respectively. At long times the distance along the vertical z between the J and S_s points is smaller than the distance between the J and S_w points. The S_w saddle points are farther from the J jet apices in the downward direction along the z axis than are the S_s saddle points. Therefore, the S_w saddle points are called ''weak'' in comparison to the ''strong'' S_s saddle points.

Hexagonal-lattice bubble array. Let us move on to the potential (3.2). First, let the amplitude a_{10} dominate (let the disturbances of η be equal to zero) at $t=0$, and, second, let $a_{10}>0$. Then it is easy to see that there are bubble apices at the B points in Fig. 2. The first few lattice points are numbered 1, 2, 3, and 4. Point 1 is the coordinate center. At the J points there are jet apices. Several jets are numbered (5–7). Symmetric verticals ($V_{\perp}=0$) pass through the bubble apices B, the jet apices J, and the saddle points S. Symmetry planes pass through the straight lines which are continuations of segments BJ, BS, and JS. The minimal domain is triangle BJS. The total flow is the sum of the parquet of domains.

Triangular-lattice bubble array. Complementariness of triangular and hexagonal lattices. Let us consider the case of a triangular-lattice bubble array (Fig. 3). We denote the corresponding potential by $\varphi_3(x_3,y,z,t)$. The case of a honeycomb array (Fig. 2) with the potential $\varphi_6(x_6,y,z,t)$, which is given by the double sum (3.2) when $a_{10}>0$, was considered above. In that case the apex B of a hexagonal bubble is located at the center ($x_6=0, y=0$). A triangular-lattice bubble array is also given by the sum (3.2), but when $a_{10}<0$. The potential $\varphi_3(x_3,\dots)$ of a triangular-lattice bubble array with a bubble at the coordinate center ($x_3=0, y=0$) is obtained from $\varphi_6(x_6,\dots)$ by displacement by the segment connecting points 1 and 5 in Fig. 2 ($x_6=4\pi/3+x_3$).

The complementariness of hexagonal and triangular bubbles follows from the invariance of the potential (3.2) with respect to rotation about the $p6$ axis (by 60°), which is located at point 1, and about the $p3$ axis (by 180°), which is located at point 5 in Figs. 2 and 3.

4. PARABOLIC MODEL. TWO-DIMENSIONAL PHASE SPACE

4.1. The potentials φ_6, φ_4, and φ_3. In model 1 the expansions (3.1) and (3.2) are cut off at the first terms. The potentials of hexagonal, square, and triangular bubbles have the form

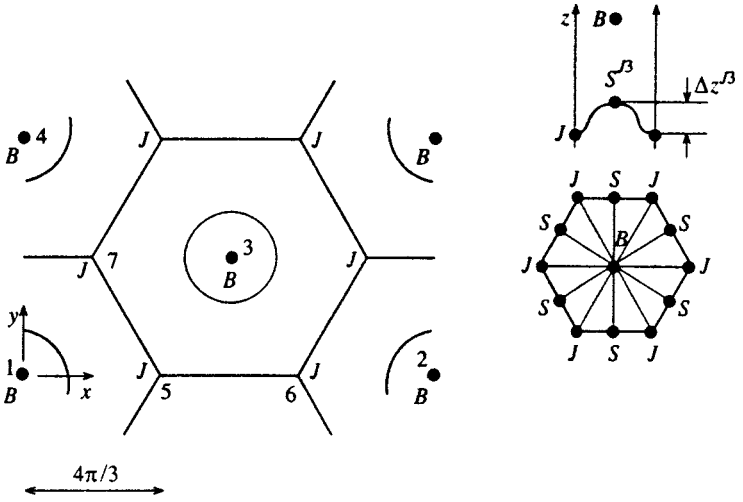

FIG. 2. Hexagonal array of bubbles B. The apices B are surrounded by circles. It is important that three symmetry planes converge at the jet apices J (see the end of Sec. 5.2).

$$\varphi_6=-\frac{a}{3}(c+c^+ +c^-)e^{-z},\quad c=\cos x,\quad c^\pm=\cos\xi^\pm, \tag{4.1}$$

$$\varphi_4=-\frac{a}{2}(\cos x+\cos y)e^{-z}, \tag{4.2}$$

$$\varphi_3=\frac{a}{6}(c-\sqrt{3}s+c^+ +\sqrt{3}s^+ +c^- +\sqrt{3}s^-)e^{-z}, \tag{4.3}$$

$$s=\sin x,\quad s^\pm=\sin\xi^\pm.$$

The bubbles are located at the center where $x=y=0$. In (4.1)–(4.3) $a=a_{10}$. According to Sec. 3, the amplitudes $a>0$ in (4.1) and (4.2), and $a<0$ in (4.3).

4.2. Fluid boundary. It can be shown that in the case of a hexagon, a square, and a triangle the principal curvatures of the surface η at B and J points coincide.[5)] Therefore, the surface η can be approximated by a symmetric paraboloid in the vicinity of apices. Accordingly, horizontal sections of the boundary η near apices have the form of small circles (see Figs. 2 and 3).[6)] In these regions we have

$$\eta(x,y,t)=\eta_0(t)-K(t)\frac{\Delta}{2},\quad \Delta=x^2+y^2. \tag{4.4}$$

4.3. Calculation of velocities and derivatives. We calculate the acceleration potential φ_t, the velocities u, v, and w, and the derivatives η_x and η_y using the direct differentiation of (4.1)–(4.4). We consider the value of the potential and the velocities at the boundary $\varphi_t|$, $u|$, $v|$, and $w|$. We calculate $u^2|$, $v^2|$, and $w^2|$. We expand the expressions in the small parameter Δ (4.4), retaining the zeroth and first terms with respect to the powers of Δ.

4.4. Laboratory and comoving coordinate frames and zeroth order of the kinematic condition. The multiplier $\dot{a}\exp(-\eta_0)$ appears during the calculation of $\varphi_t|$. Let us consider the laboratory coordinate frame and the coordinate frame comoving with a bubble apex. In the laboratory coordinate frame the fluid is at rest at infinity, and in the comoving frame the bubble apex is at rest. We introduce the important notation

$$A=a\exp(-\eta_0). \tag{4.5}$$

It is not difficult to see that A is the Fourier amplitude in the comoving coordinate frame, while a is the Fourier amplitude in the laboratory coordinate frame. Differentiating (4.5), we find

$$\dot{a}\exp(-\eta_0)=\dot{A}+\dot{\eta}_0 A. \tag{4.6}$$

We write the kinematic condition (2.1) at a B point. This gives the zeroth order of the expansion (2.1) with respect to Δ at the point where $x=y=0$. From (2.1) and (4.4) we have

$$\eta_t(x=0,y=0,t)=\varphi_z[x=0,y=0,z=\eta(0,0,t),t],$$

$$\eta_t(0,0,t)=\dot{\eta}_0(t).$$

Plugging in the velocity $w|$ calculated in Sec. 4.3, we find

$$(\dot{\eta}_0)_6=A, \tag{4.7}$$

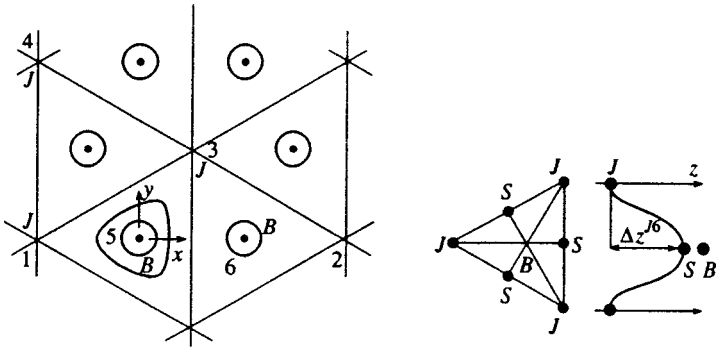

FIG. 3. Triangular array of bubbles B (circles). Six symmetry planes converge at the J apices.

$$(\dot\eta_0)_4 = A, \tag{4.8}$$

$$(\dot\eta_0)_3 = -A/2, \tag{4.9}$$

where formulas (4.7)–(4.9) refer to the cases of a hexagon ($A>0$), a square ($A>0$), and a triangle ($A<0$), respectively. Substituting (4.7)–(4.9) into (4.6), we find

$$[\dot a \exp(-\eta_0)]_6 = \dot A + A^2,$$

$$[\dot a \exp(-\eta_0)]_4 = \dot A + A^2, \quad [\dot a \exp(-\eta_0)]_3 = \dot A - A^2/2.$$

4.5. Substituting into the boundary conditions. We write out the potentials, velocities, squares of the velocities, and derivatives. The acceleration potential on the boundary is given by the expressions

$$(\varphi_t)_6 = -\dot a \frac{c + c^+ + c^-}{3} e^{-z},$$

$$(\varphi_t)_6| = -\dot a \exp(-\eta_0) \frac{c + c^+ + c^-}{3} e^{K\Delta/2} = -(\dot A + A^2) \times \left(1 - \frac{1-2K}{4}\Delta\right) \tag{4.10}$$

in the case of a hexagon,

$$(\varphi_t)_4 = -\dot a \frac{\cos x + \cos y}{2} e^{-z},$$

$$(\varphi_t)_4| = -(\dot A + A^2)\left(1 - \frac{1-2K}{4}\Delta\right) \tag{4.11}$$

in the case of a square, and

$$(\varphi_t)_3 = \dot a \frac{c - \sqrt{3}s + c^+ + \sqrt{3}s^+ + c^- + \sqrt{3}s^-}{6} e^{-z},$$

$$(\varphi_t)_3| = \left(\dot A - \frac{A^2}{2}\right)\frac{c - \sqrt{3}s + \ldots}{6} e^{K\Delta/2} = \left(\frac{\dot A}{2} - \frac{A^2}{4}\right)\left(1 - \frac{1-2K}{4}\Delta\right) \tag{4.12}$$

in the case of a triangle.

Similarly, we obtain

$$(\varphi_z)_6| = A\left(1 - \frac{1-2K}{4}\Delta\right), \tag{4.13}$$

the expression for $(\varphi_z)_4|$ coincides with (4.13), and

$$(\varphi_z)_3| = -\frac{A}{2}\left(1 - \frac{1-2K}{4}\Delta\right), \tag{4.14}$$

$$(\varphi_x)_6| = (\varphi_x)_4| = Ax/2, \quad (\varphi_y)_6| = (\varphi_y)_4| = Ay/2, \tag{4.15}$$

$$(\varphi_x)_3| = -Ax/4, \quad (\varphi_y)_3| = -Ay/4. \tag{4.16}$$

Substituting formulas (4.10)–(4.16) into the boundary conditions (2.1) and (2.2), we arrive at the dynamic system sought for the parabolic model. The system of equations has the form

$$\dot K = \frac{1-4K}{2} W, \quad \dot W = -\frac{W^2 - 4gK}{2(1-2K)}. \tag{4.17}$$

In the system (4.17) $W = \dot\eta_0$ is the bubble rise velocity, $K = 1/R$, R is the radius of curvature, $g=0$ for Richtmyer–Meshkov instability, $g=1$ for Rayleigh–Taylor instability, and $g=-1$ in the case of gravity waves (see Ref. 15). In the cases of a hexagon, a square, and a triangle, we have $W = A$, A, and $-A/2$, respectively [see formulas (4.7)–(4.9)].

We see that the system (4.17) is a universal system, which describes hexagonal-, square-, and triangular-lattice bubble arrays.

4.6. Complete analysis of the system. The phase space of the system (4.17) in the variables W and K is very simple.[15,18] In the case of 3D standing gravity waves ($g=-1$), the trajectory $W(t), K(t)$ in the W,K plane is a closed contour. Passage around the contour corresponds to one wave oscillation period. It is interesting that the lattice of wave crests is hexagonal during one half of the period and triangular during the other half.

In the cases of Rayleigh–Taylor and Richtmyer–Meshkov instability, the phase layout is exhausted by one stationary point, which is a node. All trajectories having physical meaning end at this node at $t\to\infty$. Hence follows the stability of the solutions in the trajectory class of model 1 (Sec. 1). In the stationary state we have

$$K = 1/4, \quad R_d = 4/k, \quad W = 1, \quad W_d = \sqrt{g/k} \tag{4.18}$$

in the case of Rayleigh–Taylor instability[7)] (the subscript d denotes dimensional values) and

$$K = 1/4, \quad R_d = 4/k, \quad W = 1/t, \quad W_d = 1/kt \tag{4.19}$$

in the case of Richtmyer–Meshkov instability.

The behavior of the system (4.17) for Richtmyer–Meshkov instability is especially simple. When $g=0$, it becomes homogeneous and is easily integrated. While the trajectories for Rayleigh–Taylor instability depend significantly on the initial data, in the case of Richtmyer–Meshkov instability there is only one independent trajectory. The remaining trajectories, which pertain to different initial data, are obtained from it by simple scaling. The exact integrals of the system (4.17) for $g=0$ have the form

$$\sqrt{\frac{1-2K}{1-4K}} - 1 - \frac{1}{\sqrt{2}} \ln \frac{\sqrt{2(1-2K)} + \sqrt{1-4K}}{\sqrt{2}+1} = W_0 t, \tag{4.20}$$

$$\frac{W_0}{W} - 1 + \frac{1}{2\sqrt{2}} \ln\left(\frac{\sqrt{2}+1}{\sqrt{2}-1}\,\frac{\sqrt{2}W_0 - W}{\sqrt{2}W_0 + W}\right) = W_0 t. \tag{4.21}$$

The displacement of a bubble is easily calculated from Eq. (4.21):

$$\eta_0(t) = \int_0^t \dot\eta_0(\tau)\,d\tau = \int W(\tau)\,d\tau.$$

In Ref. 18 the system (4.17), the stationary states (4.18) and (4.19), and the solutions (4.20) and (4.21) were obtained for the case of a square lattice. Here this approach has been extended to two more important lattices. The system (4.17) and the solutions (4.20) and (4.21) describe the transition from a linear to a nonlinear stationary state.

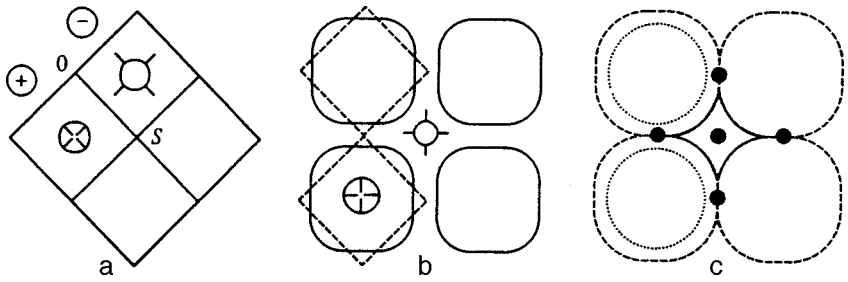

FIG. 4. Square lattice. The $h=0$ and $z=h_s$ levels in the linear (a) and nonlinear (b, c) stages. a) An $h<0$ level, which is marked by a circle with dashes directed outward, is in the square near the minus sign. The straight lines depict the separatrix network. It is marked by a zero, since the height in it is $h=0$. An $h>0$ level, which is marked by a circle with dashes directed inward, is in the square near the plus sign. b) Horizontal sections at $h=0$ in the linear [dotted lines, transferred from Fig. (a)] and nonlinear (solid curves without dashes) stages. c) Horizontal sections at $h=0$ [dotted lines, transferred from Fig. (b)] and $z=h_s$ (dashed and solid curves) in the nonlinear stage. The separatrix network consists of four-pointed stars with sharpened tips joined together at S points. One such star is depicted by solid curves.

5. BOUNDARY SURFACE RELIEF

5.1. Squares. Let us systematically examine square ($4m$), hexagonal ($6m$, Sec. 5.2), and rectangular ($2m$, Sec. 5.3) lattices. We begin with the $4m$ case. Let the potential $\varphi_4(x,y,z,t=0)$ be given by the sum of two waves (4.2), and let $\eta(x,y,0)\equiv 0$. The evolution of the spatial structure of the surface η from the linear to the nonlinear stage is traced in Fig. 4. We can describe the relief of the surface $\eta(x,y,t_{\rm obs})$ with its ''mounds'' and ''pits.'' In the linear stage [$\eta\propto w|\approx w(x,y,z=0,t)$] the shapes of the bubbles (''mounds'') and jets (''pits'') are identical.

Let us discuss the topography of horizontal sections. Consider the curves (level lines), along which the surface $z=\eta(x,y,t)$ and the horizontal plane $z=h$, where h is the height of the horizontal section of the relief, intersect. Consider also the level map. There are nonempty intersections at $h_J<h<h_B$, where $h_B=\eta(2\pi n,2\pi m,t)>0$, $h_J=\eta[(2n+1)\pi,(2m+1)\pi,t]<0$, and $n,m=0,\pm 1,\ldots$. The levels with $h>h_J+\epsilon$ $(0<\epsilon\ll 1)$ are closed around a J point, and the levels with $h<h_B-\epsilon$ are closed around a B point. One such contour closed around a J point and one such contour closed around a B point are depicted in Fig. 4(a). The dashes indicate the direction of increasing h.

In the linear stage the bubbles and jets divide the x,y plane into equal squares [a checkerboard, Fig. 4(a)]. On the black squares of the board there are jets (the dashes point outward), and on the white squares there are bubbles (the dashes point inward). As the height h increases, we have $J\rightarrow S\rightarrow B$. The contours around a bubble apex h_B give way to contours around a jet apex h_J upon passage through the separatrix network. Its points are saddle points (S). The network is formed by separatrices joining S points.

In the linear stage the separatrices are rectilinear and are all located at the zero level $h=h_0=0$. This is a manifestation of the ''degeneracy'' of the square case in the linear stage. We note that $4m$ structures are the simplest among the spatial structures.

In the nonlinear stage the coincidence of the separatrices with the h_0 level is eliminated. There is no such coincidence in either the linear stage or (for sure) in the nonlinear stage in the case of $6m$, $3m$, and $2m$ lattices (see Secs 5.2 and 5.3). In $6m$, $4m$, and $3m$ lattices there is one type of saddle point (see Figs. 2, 4, and 3). In $2m$ lattices there are two types of saddle points: S_s and S_w (see Fig. 1). Therefore, in this case the separatrix network is usually divided into two systems of parallel chains: S_s chains (segments of S_s-S_s separatrices and S_s saddle points) and S_w chains (see Sec. 5.3).

The nonlinear stage is shown in Figs. 4(b) and (c). In it the shapes of the bubbles and jets are different. The zero level h_0 is deformed. It transforms from the square depicted by solid straight lines in Fig. 4(a) and by dotted lines in Fig. 4(b) into the rounded contour depicted in Fig. 4(b) by a solid curve and in Fig. 4(c) by dots. Within this contour the height $h>0$. This ''positive'' area S^+ is the region where the ''vacuum'' rises. Outside the contour $h<0$. This is the ''negative'' area S^-, where the fluid drains downward. In the linear stage $S^-/S^+=1$. In the nonlinear stage $S^-/S^+<1$.

Let us consider the limit $t\rightarrow\infty$. A bubble asymptotically escapes to infinity from the initial position of the boundary ($h_B-h_0\rightarrow\infty$). In the process, $S^-/S^+\rightarrow 0$, and the contour h_0 is pressed against square $JJJJ$ (see Fig. 1 for $q=1$). Square $JJJJ$ is a cross section of a pipe, along whose walls flattened or wall jets travel. We see that rotation of the square by 45° takes place. At $t=0$ the $h=0$ squares are obliquely oriented [see Fig. 4(a)]. Their sides are parallel to the bisectors of the angles between the x and y axes. At $t=\infty$ the sides of the $h=0$ squares are parallel to the x and y axes.

In the nonlinear stage the $S\rightarrow S$ separatrices lose their rectilinear character. They are depicted in Figs. 4(b) and 4(c) by solid (the central star) and dashed curves. The points on the tips of the stars are saddle points (S). The center of each star is a jet apex (J). With the passage of time the separatrix intersection angle becomes increasingly acute. At $t\rightarrow\infty$, $h_s\rightarrow-\infty$ (the density ratio $\mu=0$). The star then contracts into a cross of JS segments.

The characteristic levels form the sequence $h_J<h_s<h_0<h_B$. The contour of the horizontal section at $h=0$ [dotted curves, Fig. 4(c)] is located within the h_s contour (dashed and solid curves).

Let us compare Fig. 4 and Fig. 1 in the case of $q=1$, where the S_s and S_w saddle points are identical. Figures 1 and 4 show downward views (in the x,y plane). In addition, Fig. 1 presents a lateral view (in a plane passing through the

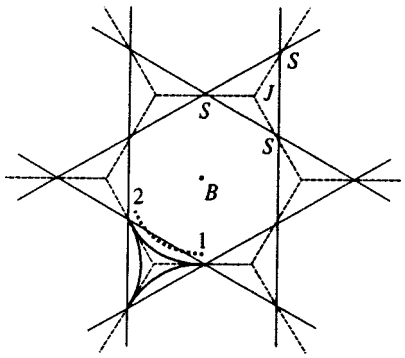

FIG. 5. Level of the hexagonal surface η. In the linear stage the $h<h_s$ region is bounded by triangle SSS, which is composed of rectilinear separatrices. In the nonlinear stage the separatrices become curves and the triangle transforms into a three-pointed star. A group of six stars surrounds each bubble B.

z axis), which is important for understanding the spatial structure of the flow. The lateral view of the flow shows the J and S points of the dowarnward moving arch JSJ and the projection of an upward moving B point onto a JJ plane. The JJ planes are symmetry planes or lateral walls of pipes confining a flow cell. Arc JS (a planar curve; half of arch JSJ) lies in the line of intersection of a wall jet and a JJ plane. It is the lower edge of the wetted part of the pipe wall. In the horizontal view it is represented by segment JS.

For the spatial orientation we offer two variants of rise from a J point to a B point on the surface η. In the first variant rise occurs along arch JB, whose projection onto the x,y plane is segment JB. In the second variant rise occurs in two stages. In the first stage we rise from J to S along arc JS, and in the second stage we pass from S to B along arc SB. The projections of these arcs onto the x,y plane are perpendicular to one another (see Fig. 1). Together, these segments form triangle JBS.

5.2. Hexagons. Section 7 contains a comparative analysis of $6m$, $4m$, and $3m$ lattices. It is found that the $6m$ case is unique.[8)] For this reason we shall dwell on it.

Let $\eta(x,y,0)\equiv 0$, and let the potential $\varphi_6(x,y,z,0)$ be given by the sum of three waves (4.1). As in Sec. 5.1, let us consider the characteristic levels h_J, h_s, h_0, and h_B. In the linear and nonlinear stages we have $h_s<h_0$. Using the trigonometric formulas for cosine addition, it is not difficult to show that in the linear stage, in which $h\propto w|$ and the velocity w is determined only by the linear combination of harmonics (4.1), the network of $S-S$ separatrices consists of straight lines. They are depicted by solid lines in Fig. 5. The network is formed from equilateral SSS triangles with centers at J points. Six triangles are linked around the apex B.

In the linear stage the $S-S$ separatrices are curved. Three-pointed stars with sharpened tips form around the J apices. At $t\to\infty$ the $h=0$ level, one-sixth of whose contour at $t\neq\infty$ is depicted by arc 1–2 in Fig. 5, is pressed against the walls of the pipe composed of JJ planes. They are depicted by dashed straight lines in Fig. 5. At $t=\infty$ the rays degenerate into JS segments.

Jets drain along the JJ walls. Because of the spreading in the plane of a wall, they are flattened (flattened, wall, or near-wall jets). The jets are stronger at the corners of the pipe. Because of this, the indentation depth of the J points is greatest on the JS arcs. The degree of bulging of the curved jets J from the ''wall'' is characterized by the ratio

$$\zeta=(h_J-h_s)/h_J. \tag{5.1}$$

It is small in the case of wall jets. In the case of a finger-like jet J this ratio is large.

Let us compare the topology of the jets in $6m$ (Figs. 2 and 5) and $3m$ (Fig. 3) lattices. In a $6m$ lattice the bulging of the curved jets J from the the walls JS is weak, and the ratio (5.1) is small. Accordingly, the bubbles in a $6m$ lattice are well isolated from one another by the walls JS separating them along the entire bubble perimeter.

In a $3m$ lattice the situation is reversed. The bubbles are poorly isolated from another. The walls separating them along the BSB directions through a saddle point are relatively low, and, for this reason, the bulging of the curved jets J is significantly greater. Therefore, the ratio ζ (5.1) is fairly high. An array of strong finger-like jets J forms.

This is the source of the specific features of a $3m$ lattice. Enhanced focusing of the jets occurs in it. Jets are generated by bubbles. In a $3m$ lattice the six bubbles arranged in a circle about one jet J ''feed'' that jet and make it strong.

Let us compare the types of rise from a J point to a B point through an S saddle point. In the $6m$ case the difference between the heights of the B and J points equals $(z_B-z_s)+\Delta z^{J3}$. In the $3m$ case this difference equals $(z_B-z_s)+\Delta z^{J6}$, where the segments Δz^{J3} and Δz^{J6} are defined in Figs. 2 and 3, and the numbers 3 and 6 in the superscripts $J3$ and $J6$ indicate how many symmetry planes converge at a J point. The difference Δz^{J3} is smaller than Δz^{J6}. Therefore, as has been stated, in a $6m$ lattice the jets are of the wall type, and in the $3m$ case a system of strong finger-like jets forms. It is shown in Sec. 7 that the intersection of more than three wall jets is unlikely. Therefore, in 3D flows the type of three-pointed star shown in Fig. 5 (see also Fig. 2) is the principal type for the jets.

5.3. Rectangles. The structure of $2m$ lattices is significantly more complicated than the high-symmetry cases considered above. This is caused by the splitting of the saddle points into two types and the increase in the number of characteristic points. Because of the splitting, the network separatrix diagrams give way to striped diagrams.

Let $\eta(x,y,0)\equiv 0$, and let the potential at the initial moment be given by a sum of two harmonics and have the form

$$\varphi(x,y,z,t=0)=-\frac{a(0)}{2}\cos x e^{-z}-\frac{b(0)}{2q}\cos qy e^{-qz}, \tag{5.2}$$

where $a=a_{10}$ and $b=a_{01}$ [compare this with formulas (3.1) and (4.2)]. To fix ideas, we set $a(0)/b(0)=1$ and $q>1$. The structure of the surface η at $t>0$ is shown in Fig. 6. Figure 6 has been stretched in the y direction by a factor of q (compare it with Fig. 1).

The laws governing the growth of disturbances with time are different in the cases of Rayleigh–Taylor and Richtmyer–Meshkov instability. Let us consider them separately. We begin with Rayleigh–Taylor instability. It is sig-

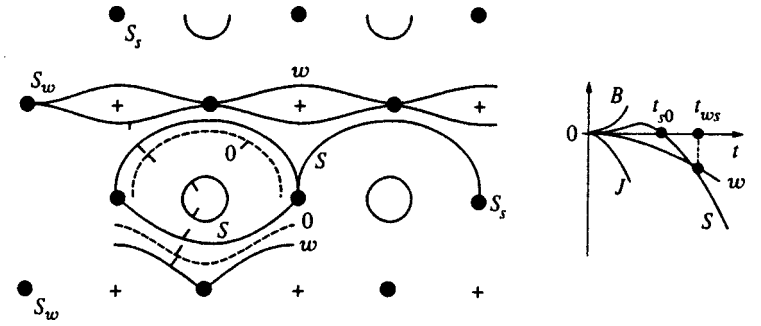

FIG. 6. Level map of η in a $2m$ lattice. Chain-like structure of S_s-S_s and S_w-S_w separatrices. There are no S_s-S_w separatrices, which would have joined the chains in a network. Circles—bubbles, crosses—jet apices, 0—zero level, s—S_s-S_s separatrices, w—S_w-S_w separatrices. The dashes on the levels are drawn in the direction of growth of h. Right-hand figure—dependence of the positions of characteristic points for Rayleigh–Taylor instability on t.

nificant that the wavelengths of the harmonics in (5.2) are different. This distinguishes the potential of a $2m$ lattice from the more symmetric potentials (4.1)–(4.3). In the linear stage a wave grows more rapidly in the y direction, since $a(t)=a(0)e^t$ and $b(t)=b(0)e^{\sqrt{q}t}$. We recall that instability normalization to $k=1$ and $g=1$ is chosen in the case of Rayleigh–Taylor and that $2\varphi(x,y,z,t)=-a(t)c_x e^{-z}-b(t)c_{qy}e^{-qz}$. Therefore, a separatrix diagram with cells that would connect the S_s and S_w saddle points in a network is lacking already in the linear stage. Instead of a network, separatrix chains of types w (S_w-S_w) and s (S_s-S_s) appear (see Fig. 6).

During motion the positions of the S_s points (curve s in Fig. 6) and the S_w points (curve w) relative to one another and relative to the $z=0$ level vary. The hierarchies of heights have the following forms: $h_{sw}<0<h_{ss}$ at $0<t<t_{s0}$, $h_{sw}<h_{ss}<0$ at $t_{s0}<t<t_{ws}$, and $h_{ss}<h_{sw}<0$ at $t>t_{ws}$, where the subscripts sw and ss refer to S_w and S_s saddle points. The lower half of Fig. 6 shows the $h_{sw}<h_0<h_{ss}$ levels (curves w, 0, and s), which refer to the stage at $0<t<t_{s0}$. The subsequent stage, in which these levels are arranged in the order of 0, s, and w, is depicted in the upper half.

If the areas in the x,y plane within which $h<h_{sw}$ are hatched, a system of separated stripes stretching parallel to the x axis is obtained. In the final stage at $t>t_{ws}$ the S_s level becomes lower. The orientation of the stripes then changes from parallel to the x axis to parallel to the y axis.

In the final stage the motion associated with the longer side of the rectangle dominates (see Fig. 1; since $q>1$, this side is x). The main flow of fluid occurs in the near-wall jets parallel to the y axis, which drain along the JS_sJ walls. In this stage $R_x>R_y$, where R_x and R_y are the radii of curvature at the apex B along the x and y axes. Let us transfer the contour of the h_0 level from the upper half of Fig. 6 to the rectangle $JJJJ$ in Fig. 1. We use Δ_x and Δ_y to denote the distances from this level to the S_w and S_s points, respectively. The gaps $2\Delta_x$ and $2\Delta_y$ characterize the thickness of the near-wall jets traveling along the JS_wJ and JS_sJ walls. The predominance of the JS_sJ jets means that $\Delta_x<\Delta_y$.

An analysis of rectangular lattices clearly reveals how the flow changes upon the transition from the 3D case to the 2D case. As $q\to\infty$, a rectangle becomes similar to a slot (Fig. 1), in which the transverse components of the motion are directed mainly along the x axis.

Let us examine Richtmyer–Meshkov instability. If the difference between $a(t)$ and $b(t)$ in the initial stage is neglected, the separatrix diagram takes the form of a network. It is similar to the one shown in Fig. 4(a). In the 2D geometry the jet apex velocity $\dot{\eta}_J$ increases from the initial value $a(0)$ to a value roughly two times larger. It is not entirely clear how the competition between $h_{sw}(t)$ and $h_{ss}(t)$ takes place in Richtmyer–Meshkov instability. In any case these functions are not equal, and a striped structure forms instead of an S_s-S_w network. We stress that the removal of mass by jets occurs mainly in the JS_sJ jets, as in the case of Rayleigh–Taylor instability.

6. COMPARISON OF THE THEORY WITH A NUMERICAL SIMULATION

6.1. Numerical method. The complete system of Euler equations for a compressible inviscid medium written in the divergence form[37,38] was integrated. A quasimonotonic grid-characteristic scheme for a second order of approximation was used. The monotonicity was enhanced by a combination of schemes with central and oriented differences. A similar hybrid method was developed for numerically simulating flows of an incompressible fluid.[39] No artificial viscosity, no smoothing, and no flow-constraining procedures were used in the calculation scheme. The scheme has such useful qualities as conservatism, monotonicity, and an increased approximation order. The requirement for monotonicity ensures nonlinear dissipation, which smooths any short-lived perturbations with a wavelength of the order of several grid steps. The integration region has the form of a rectangular parallelepiped, on whose lateral sides the symmetry conditions are satisfied and on whose lower and upper faces the zero-flow conditions are satisfied.

6.2. Rayleigh–Taylor instability. Figures 7–9 present the results of the calculation of Rayleigh–Taylor instability. A quarter of the total flow period in a $\pi\times\pi\times 7\pi$ cell was calculated on a $30\times 30\times 210$ grid. Figure 7 shows a complete period. The initial position of the boundary coincides with the $z=0$ plane (Fig. 7), which divides the cell along its height in a 3:4 ratio. The ratio between the densities of the lower and upper fluids is $\mu=1/10$, and the normalization is to $g=1$ and $k=1$. The initial disturbance is given by Eq. (4.2) with $a(0)=0.05$.

The displacement of a bubble $\eta_B(t)$ is shown in Fig. 8. It is seen that a regime with a constant bubble rise velocity is established. In the numerical experiment the limiting velocity is $w_{cs}(\infty)=0.83$, which amounts to roughly 90% of the theoretical rate (4.18) corrected by the factor $\sqrt{1-\mu}$, which takes into account the finite density of a light fluid.[6,40] Figure

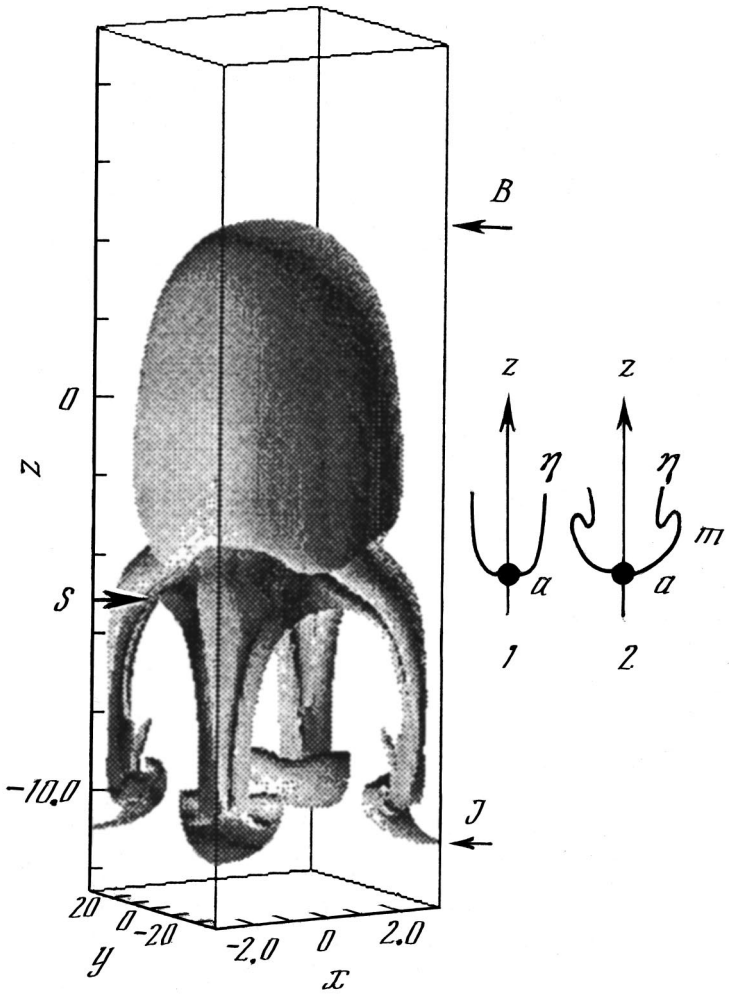

FIG. 7. Form of the interface in the case of a square-lattice bubble array at the time $t=9$. The arrows mark S saddle points and B and J apices. Right-hand figure—scheme for the appearance of a mushroom formation.

9 shows theoretical and numerical plots of the radius of curvature $R(t)=1/K(t)$. As we see, the theory and simulation are in satisfactory agreement.

6.3. Richtmyer–Meshkov instability. Figures 10 and 11 show the results in the case of Richtmyer–Meshkov instability ($g=0$). The geometric dimensions and the ratio μ were the same as in Sec. 6.2. The initial perturbation was given by Eq. (4.2) with $a(0)=1$. As we see, the results with respect to the velocity (and thus with respect to the displacement) are in good agreement. The descriptions of the bubble surface bending process [the plots of $K(t)$] are in satisfactory agreement. The bubble curvature in the numerical simulation is somewhat greater than the calculated curvature at large times.

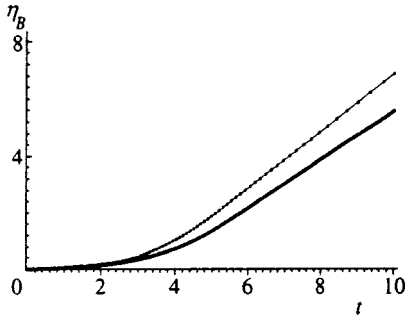

FIG. 8. Plot of $\eta_B(t)$. Thin curve with markers—theory in Sec. 4, thick curve—simulation.

6.4. Role of the density ratio. The theory developed in Secs. 4 and 5 refers to the case of $\mu=0$. Let us examine the changes occurring when $\mu \ll 1$. The shape of the bubbles scarcely changes.[18] Their rate of motion varies slightly.[6,40] At the same time, there is a qualitative detail, which vanishes when $\mu=0$. It is the appearance of mushroom-shaped formations.

In the case of Rayleigh–Taylor instability, the falling matter accelerates to high speeds in the jets (therefore, the jets are narrow). As a result, velocities are achieved at which the influence of aerodynamic drag on the part of the light fluid on the advancement of the jet apex becomes important. The aerodynamic thrust of the light fluid leads to the formation of mushrooms. This question was investigated in Ref. 40 for the case of the 2D geometry using conformal mappings and a hodograph technique. The thrust of the light fluid is also the cause of the formation of mushroom-shaped jets in Richtmyer–Meshkov instability.

The formation of mushrooms is a consequence of the stagnation of the heavy fluid due to the thrust of the light one. The stagnation zone is localized near the point of separation of the flow of the light fluid by the jet of the heavy

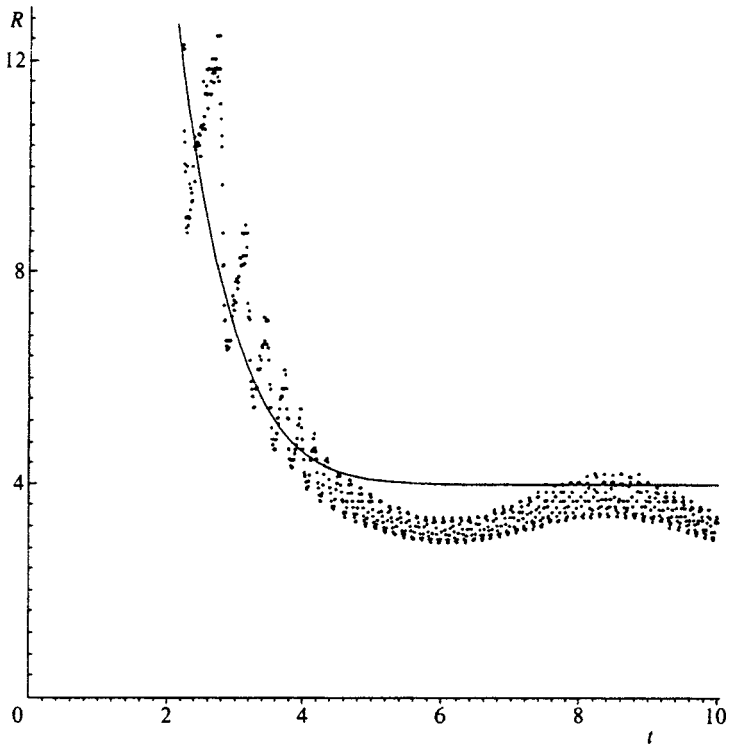

FIG. 9. Plot of $R(t)$. Solid curve—calculation of the system (4.17), markers—numerical simulation.

light.[40] The jet apex is the separation point. A mushroom forms when the jet becomes sufficiently thin and gains a sufficiently high speed. The jet is thin where the radius of curvature R is small. When $\mu=0$, R is smallest at the jet apex.

Matters are the same in the 3D geometry. Now the ''thin fast'' zone is a stripe along a JSJ arch (see Sec. 5). Here the rate of motion of the light fluid against the arch is high, and the radius of curvature of the surface η in the direction perpendicular to the plane of the arch is small. Separation of the flow of the light fluid occurs on the arch. Its streamlining past the arch leads to the formation of a stripe-like mushroom-shaped formation along the entire arch.

Let us consider a section of the arch and of the near-wall jet following it in plane P, which is perpendicular to the arch. It is shown on the right in Fig. 7. Here a is the point of intersection of plane P and arch JSJ, $a\eta$ is the curve of intersection of plane P and the surface η, and az is the plane of the wall along which the near-wall jet drains. Section 1 refers to the case of $\mu=0$, and section 2 refers to the case of $\mu>0$. When $\mu>0$, the mushroom m forms, as occurs in the 2D geometry.[40]

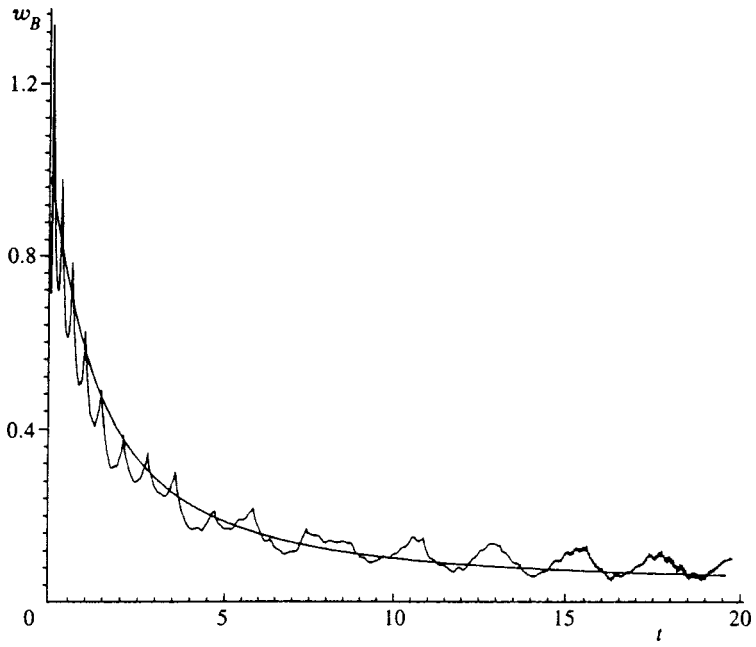

FIG. 10. Plots of the bubble velocity $w_B(t)$. Smooth curve—the integral (4.21), fluctuating curve—simulation.

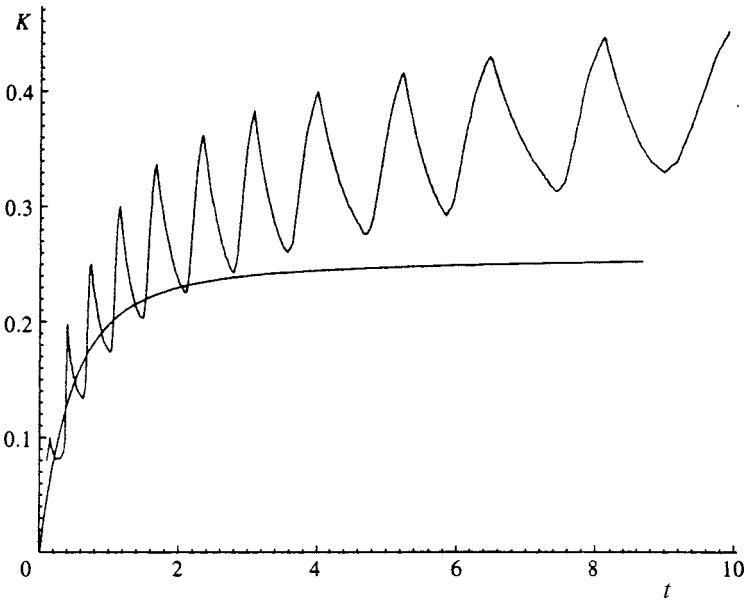

FIG. 11. Plots of the bubble curvature $K(t)$. Smooth curve—the integral (4.20), oscillating curve—simulation.

The arc of the mushroom head am cuts a groove along arch JSJ as point a moves along the arch. In the case of a $4m$ lattice four such grooves descend toward each apex J from four sides. In a horizontal section these grooves are located near the tips of the four-pointed star shown in Fig. 4(c). The grooves end near a toroidal mushroom, which crowns the apex J. Thus, the situation is fairly far from the picture of a jet which is axisymmetric relative to the vertical passing through the apex J.

The grooves and the torus are clearly seen in Fig. 7. The mushroom-shaped formation begins near an S saddle point and stretches toward a J apex. The grooves, as they should be, are adjacent to a wall of the ''pipe'' (a boundary of the calculation region), since the arch JSJ lies in the plane of this wall.

7. COMPARISON OF LATTICES

Let us compare the $6m$, $4m$, and $3m$ lattices. We expand their potentials in combinations of single wave vectors and their overtones (3.1) and (3.2). Then in the parabolic approximation (Sec. 4) the bubble rise velocities in the $6m$, $4m$, and $3m$ lattices are identical both in the case of Rayleigh–Taylor instability and in the case of Richtmyer–Meshkov instability. It is found for Rayleigh–Taylor instability that both the linear stage (equality of the growth rates for $k=1$) and the entire transition process [the system (4.17)] are identical in these lattices. The linear stage of Richtmyer–Meshkov instability is determined not only by the wave vectors, but also by the initial velocities. If we start from a flat boundary and choose equal initial velocities, then, as in the case of Rayleigh–Taylor instability, both the linear and nonlinear stages are identical in all three lattices. This follows from the results of Secs. 4.5 and 4.6.

Let us compare the lattices in another respect. For a fixed wave number k the areas per bubble in the x,y plane in the $6m$, $4m$, and $3m$ lattices are, respectively,

$$S_6=\frac{8\pi^2}{\sqrt{3}k^2},\quad S_4=\frac{4\pi^2}{k^2},\quad S_3=\frac{4\pi^2}{\sqrt{3}k^2}$$

(see Figs. 1–3). These areas are not equal, and $S_6:S_4:S_3=(2/\sqrt{3}):1:(1/\sqrt{3})\approx 1.15:1:0.577$. It is significant that the area of a triangular cell differs significantly from the roughly equal areas of square and hexagonal cells.

Let us require that the bubble areas be identical in the lattices being compared (the condition of an equal number of bubbles per unit area). Then the wave numbers are related by the ratio

$$k_6:k_4:k_3=\sqrt{\frac{S_6}{S_4}}:1:\sqrt{\frac{S_3}{S_4}}=\frac{\sqrt{2}}{3^{1/4}}:1:\frac{1}{3^{1/4}}\approx 1.07:1:0.76.$$

In this case $\lambda_6:\lambda_4:\lambda_3=k_6^{-1}:k_4^{-1}:k_3^{-1}\approx 0.93:1:1.3$. The radii R_6, R_4, and R_3 form the same ratio in Rayleigh–Taylor and Richtmyer–Meshkov instability. According to the rate of the transition process, the lattices are arranged in the order of 6, 4, and 3, i.e., the transition is slowest in the triangular lattice. In return, the limiting bubble velocity is highest in the $3m$ lattice. These velocities are related as

$$w_6:w_4:w_3=(S_4/S_6)^{1/4}:1:(S_4/S_3)^{1/4}=\sqrt{\lambda_6}:\sqrt{\lambda_4}:\sqrt{\lambda_3}$$
$$=(3^{1/8}/2^{1/4}):1:3^{1/8}\approx 0.97:1:1.15$$

in the case of Rayleigh–Taylor instability (4.18) and as $\lambda_6:\lambda_4:\lambda_3$ in the case of Richtmyer–Meshkov instability (4.19).

Comparing the results, we arrive at the conclusion that the $6m$ and $4m$ bubble arrays are roughly equivalent (the difference in the limiting velocity does not exceed a few percent), but the $3m$ lattice stands apart. It has the strongest system of J jets (Sec. 5.2), and the bubble velocities exceed the velocities in the other lattices by 15% in the case of Rayleigh–Taylor instability and by 30% in the case of Richtmyer–Meshkov instability.

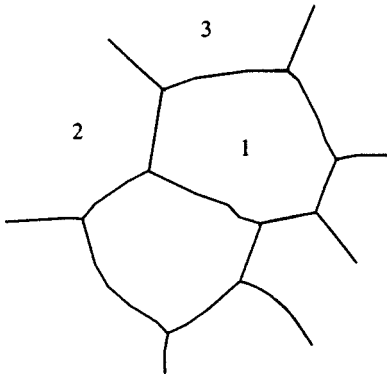

FIG. 12. Random collection of bubbles. Each bubble is bounded by wall jets. The junctions with $N_r=3$ are typical.

8. FORMATION OF THREE-FOIL JETS

The 3D structure of the surface η was considered above. Its fundamental elements are bubbles and jets. The bubbles are round and isolated, and the jets consist of extended regions (near-wall or wall jets), which form a network, and junctions, at which the wall jets intersect. The type of periodicity determines the number of wall jets N_r intersecting at a junction, and the ratio of the number of bubbles to the number of jets N_B/N_J.

Let us consider a random surface appearing in the case of 3D turbulence. Structurally, it consists of the same elements (isolated bubbles, wall and junction jets; see Fig. 12). However, the long-range order gives way to short-range order, as occurs upon the transition from a crystal to a liquid. Let us see what the number N_r equals. This would allow us to ascertain to which of the lattices the random surface is topologically closest.

It is fairly clear that the junctions with $N_r=3$ dominate. In fact, any bubble is crowded by neighbors. The adjacent bubbles are separated by a boundary, i.e., a wall jet. Its position is determined mainly by two neighbors, for example, by bubbles 1 and 2 (see Fig. 12). Let us rise along a boundary. At a certain point the influence of a third neighbor becomes important. Here the boundary which separated bubbles 1 and 2 branches. One branch passes between bubbles 1 and 3, and the other passes between bubbles 2 and 3. Therefore, the value $N_r=3$ is typical. The ratio N_B/N_J is then between 0.5 and 1, which correspond to the $6m$ and $4m$ lattices of bubbles.

9. PARABOLIC DESCRIPTION OF A RECTANGULAR LATTICE. FOUR-DIMENSIONAL PHASE SPACE

9.1. Potential and boundary. Let us apply the Layzer model to the $2m$ lattice. The potential cut off at two harmonics (3.1) [see also Sec. 5.3 and Eq. (5.2)] has the form

$$\varphi(x,y,z,t)=-\frac{\hat a(t)}{2}\cos x e^{-z}-\frac{\hat b(t)}{2q}\cos qx e^{-qz}, \quad (9.1)$$

and the expansion of the boundary at a bubble apex[9] has the form

$$\eta(x,y,t)=\eta_0(t)-\frac{Kx^2}{2}-\frac{Qy^2}{2}.$$

9.2. Expansions. We proceed as in Secs. 4.2–4.4. Now the coordinates x^2 and y^2 appear in the expressions separately, rather than in the form of the sum (4.4). Our goal is to calculate $\varphi_t|$, $u|$, $v|$, $w|$, η_x, and η_y. We present the most significant points. We differentiate (9.1) with respect to t. The calculation of φ_t on the boundary gives

$$\varphi_t|=-\frac{d\hat a}{dt}\exp(-\eta_0)\frac{\cos x}{2}e^{-\Delta\eta}-\frac{d\hat b}{dt}\times\exp(-q\eta_0)\frac{\cos qy}{2q}e^{-q\Delta\eta}, \quad (9.2)$$

where $\Delta\eta=\eta-\eta_0=-(Kx^2+Qy^2)/2$. We introduce the notations

$$A(t)=\hat a(t)\exp[-\eta_0(t)], \quad B(t)=\hat b(t)\exp[-q\eta_0(t)]. \quad (9.3)$$

Let us express the derivatives $d\hat a/dt$ and $d\hat b/dt$ in (9.2) in terms of $\dot A$ and $\dot B$. For this purpose we differentiate (9.3). We then have

$$\frac{d\hat a}{dt}\exp(-\eta_0)=\dot A+A\dot\eta_0,$$

$$\frac{d\hat b}{dt}\exp(-q\eta_0)=\dot B+qB\dot\eta_0. \quad (9.4)$$

Let us now calculate $\dot\eta_0$. For this purpose we write the kinematic condition (2.1) at an apex. We then have $\eta_t=\dot\eta_0=w|$. Plugging in (9.1), we obtain

$$\dot\eta_0=(A+B)/2. \quad (9.5)$$

Plugging (9.5) into (9.4), we find

$$\frac{d\hat a}{dt}\exp(-\eta_0)=\frac{2\dot A+A^2+AB}{2},$$

$$\frac{d\hat b}{dt}\exp(-q\eta_0)=\frac{2\dot B+qAB+qB^2}{2}. \quad (9.6)$$

Expanding (9.2) in the small variables x^2 and y^2, with consideration of (9.6) we find

$$\varphi_t|=-\frac{2\dot A+A^2+AB}{4}\frac{(K-1)x^2+Qy^2}{2}-\frac{2\dot B+qAB+qB^2}{4q}\frac{qKx^2+q(Q-q)y^2}{2}.$$

The insignificant functions of t which are homogeneous with respect to x and y have been omitted.

Similarly, we have

$$\varphi_z|=\frac{A+B}{2}+\frac{(KA-A+qKB)x^2+(QA+qQB-q^2B)y^2}{4}.$$

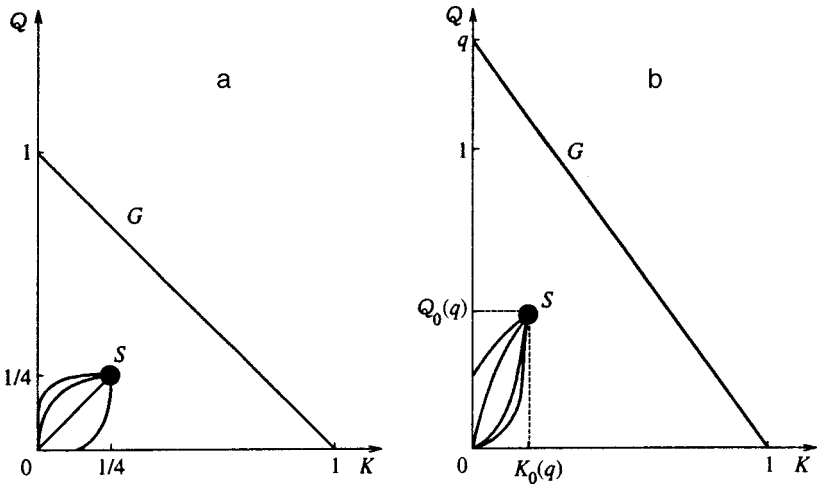

FIG. 13. Phase space and trajectories. Point S is a stationary point (node), which attracts trajectories emanating from weakly disturbed initial data: a—$q=1$, b—$q>1$.

9.3. Boundary conditions and dynamic system. Substituting the expansion from Sec. 9.2 into the boundary conditions (2.1) and (2.2), we arrive at the four-dimensional dynamic system

$$\dot{K}=(1-3K)a-qKb, \tag{9.7}$$

$$\dot{Q}=-Qa+q(q-3Q)b, \tag{9.8}$$

$$(1-K)\dot{a}-K\dot{b}=-a^2+gK, \tag{9.9}$$

$$-Q\dot{a}+(q-Q)\dot{b}=-q^2b^2+gQ \tag{9.10}$$

for unknown K and Q and the amplitudes $a=A/2$ and $b=B/2$. If $Q=0$ and $b=0$ or if $K=0$ and $a=0$, we return to the planar (2D) system,[13,15] and if $q=1$, $K=Q$, and $a=b$, we return to the system (4.17). The cross terms related to the interaction of waves along the x and y directions are qKb in (9.7) and $K\dot{b}$ in (9.9).

We note the $x\leftrightarrow y$ symmetry. Let us write φ and η for an arbitrary wave number k. The potential has the form

$$\varphi=-\frac{\hat{a}}{2k}\cos kxe^{-kz}-\frac{\hat{b}}{2qk}\cos qkxe^{-qkz}.$$

Let us derive the equations of the system. After the inversion $k\to qk$, $q\to 1/q$ (rotation of the rectangle by 90°), the equations remain invariant to the replacements $K\to Q$, $Q\to K$ and $a\to b$, $b\to a$ (interchange of the axes).

The system (9.7)–(9.10) allows a complete investigation. Its phase space is similar to the phase space of the system (4.17). Again there is a single node, which captures all the trajectories having physical meaning. We shall next show this.

9.4. Rayleigh–Taylor instability. Let $g=1$, and let us find a stationary solution. We set $\dot{K}=\dot{Q}=\dot{a}=\dot{b}=0$. We eliminate the unknown $a>0$ and $b>0$ using Eqs. (9.9) and (9.10) and Q using Eqs. (9.7) and (9.8). As a result, we arrive at the equation for K. It has the form

$$8K^2-[6+(q-1)/3]K+1=0.$$

Comparing with the case of a square ($q=1$), we find that only the root

$$K_0(q)=(q+17-r)/48, \quad r=+\sqrt{q^2+34q+1} \tag{9.11}$$

has physical meaning. The remaining functions sought are given by the expressions

$$Q_0(q)=\frac{q-1}{3}+K_0(q), \quad a_0(q)=+\sqrt{K_0(q)},$$

$$b_0(q)=+\frac{\sqrt{Q_0(q)}}{q}. \tag{9.12}$$

The stationary state defined by (9.11) and (9.12) is the only one in the region having physical meaning.

Let us investigate the stability.[10] We linearize the system (9.7)–(9.10) near the stationary state:

$$K=K_0+\delta Ke^{\lambda t}, \quad Q=Q_0+\delta Qe^{\lambda t},$$

$$a=a_0+\delta ae^{\lambda t}, \quad b=b_0+\delta be^{\lambda t}.$$

The linearization leads to the matrix

$$\begin{matrix}
-\lambda-3a_0-qb_0 & 0 & 1-3K_0 & -qK_0 \\
0 & -\lambda-a_0-3qb_0 & -Q_0 & q^2-3qQ_0 \\
1 & 0 & -(1-K_0)\lambda-2a_0 & K_0\lambda \\
0 & 1 & Q_0\lambda & -(q-Q_0)\lambda-2q^2b_0
\end{matrix}$$

for the eigenvalues $\lambda_1(q)-\lambda_4(q)$. When $q=1$, the determinant of the matrix (the characteristic equation) has the form $2\lambda^4+14\lambda^3+35\lambda^2+36\lambda+12=0$ with the roots $\lambda_{1,2}=(-6\pm 2\sqrt{3})/4$ and $\lambda_{3,4}=-2$. The stability of the system (4.17) is obvious from an analysis of the phase plane. Linearization of Eqs. (4.17) leads to the system $(\lambda+2)\delta K=0$, $2\delta K-(\lambda+2)\delta a=0$, which has the degenerate root -2. When $q\neq 1$, the coefficients of the characteristic equation are cumbersome functions of the ratio between the sides of the cell q (Fig. 1). It follows from calculations of this equation that in the interval $0<1/q<1$ the roots obey the inequalities $\lambda_1<0$, $\lambda_2<0$, $\mathrm{Re}\lambda_3<0$, and $\mathrm{Re}\lambda_4<0$. The roots λ_3 and λ_4 are degenerate when $q=1$. They become real, different, and negative near $1/q=0$.

Let us analyze the results obtained. It was demonstrated that the stationary state S defined by (9.11) and (9.12) is stable (is a node). The projection of the four-dimensional phase space of the system (9.7)–(9.10) onto the K,Q plane is

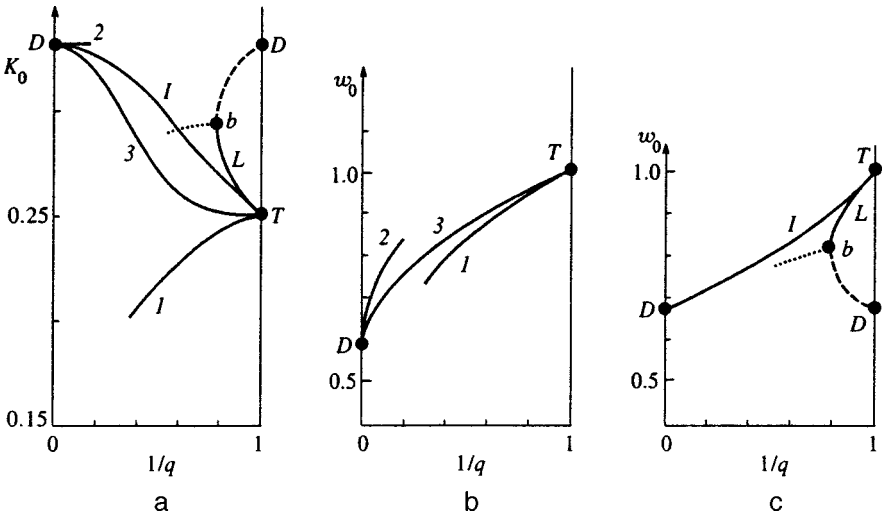

FIG. 14. Dependence of the curvature K and the velocity w on the ratio q between the sides of the rectangle. Curves *1*, *2*, and *3*—Rayleigh–Taylor instability (a and b). Curve *2* (a), point D (a and c), and curves L and I—Richtmyer–Meshkov instability (a and c).

shown in Fig. 13. Let us examine the matrix G of coefficients for the derivatives $\dot{K}$, $\dot{Q}$, $\dot{a}$, and $\dot{b}$ of this system. Its determinant is given by the expression $K+Q/q=1$. On the surface $\det G=0$, and the accelerations $\dot{a}$ and $\dot{b}$ vanish at infinity.[13,15] The projection of the surface is marked by the letter G in Fig. 13. The surface G bounds a physical region. The physical trajectories start from the center at $K=Q=0$ (if the boundary is flat at $t=0$) and its vicinity. At $t\to\infty$ they all end at the node S.

The case of $q=1$ is shown separately in Fig. 13(a). As we see, bubbles with $2m$ symmetry raise their symmetry at $t\to\infty$ and transform into bubbles with $4m$ symmetry. The trajectories passing along the bisector $K=Q$ apply to the symmetric case $[q=1,\ a(0)=b(0),$ and $K(0)=Q(0)]$ with the two-dimensional phase space of the system (4.17), $K=Q$, and $a=b$ in Fig. 13(a).

Plots of the dependence of the curvature along the long direction K ($q>1$, see Fig. 1) and of the bubble rise velocity w on $1/q$ are shown in Fig. 14. Curves *1* were calculated using formulas (9.11) and (9.12) ($w=a+b$), and curves *2* are the asymptotes for $q\gg 1$ (flow in a slot). In a slot-shaped rectangle the flow achieves a 2D regime. Motion along the long side dominates. In this case $K\approx 1/3$, $Q\approx 1/3q$, and $w\approx 1/\sqrt{3}+1/\sqrt{3q}$. The solutions (9.11) and (9.12), of course, are not suitable for small values of $1/q$ (allowance for additional harmonics with respect to the x axis is necessary). The curves which interpolate dependences *1* and *2* in the transition region are labeled by the number *3*.

9.5. Richtmyer–Meshkov instability. Let $g=0$, and let us find the stationary state and investigate its stability. In the case of Richtmyer–Meshkov instability $K=K_0$, $Q=Q_0$, $a=\alpha_0/t$, and $b=\beta_0/t$ in the stationary state.[13,15,17] Substituting these relations into the system (9.7)–(9.10), we arrive at the algebraic system

$$(1-3K_0)\alpha_0=qK_0\beta_0,\quad Q_0\alpha_0=q(q-3Q_0)\beta_0,$$

$$-(1-K_0)\alpha_0+K_0\beta_0=-\alpha_0^2,$$

$$Q_0\alpha_0-(q-Q_0)\beta_0=-q^2\beta_0^2$$

for unknown K_0, Q_0, α_0, and β_0. Successively eliminating β_0, α_0, and Q_0, we obtain the equation

$$24(3-q)K_0^3+(q^2+46q-75)K_0^2+2(13-11q)K_0+3(q-1)=0 \tag{9.13}$$

for $K_0(q)$. The correct root is selected on the basis of the condition $K_0(1)=1/4$ [see (4.19)]. Let $K_0(q)$ be a given root. The remaining functions are expressed in terms of it. We then have

$$Q_0=\frac{1-3K_0}{3-8K_0}q,\quad \alpha_0=1-\frac{1}{q}+\left(\frac{3}{q}-1\right)K_0,$$

$$\beta_0=\frac{1-3K_0}{qK_0}\alpha_0,\quad w_0=\alpha_0+\beta_0. \tag{9.14}$$

Plots of the functions $K_0(q)$ and $w_0(q)$ are presented in Fig. 14 (curves L). Curves L end at points b with $q\approx 1.26$. Two roots of Eq. (9.13) merge at that point. The real part of the solution ($\mathrm{Re}\,K_0$,$\mathrm{Re}\,w_0$) after this merging is depicted by dotted lines. The imaginary part comprises a small ($\simeq 0.1$) fraction of the real part. Curves L begin from the point T, which refers to the case of a square lattice. It is interesting that another branch (Db) emanates from point D, which corresponds to the planar (2D) solution, when $q=1$. When $q\neq 1$, it corresponds to bubbles with weak motion along the y axis. The interpolation curve I joins the asymptotes corresponding to a slot (point D) and the solution L in the intermediate region. We note that the plots of $K(q)$ referring to Rayleigh–Taylor instability (curve *3*) and Richtmyer–Meshkov instability (curve I), which have identical ends at $1/q=0$ and $q=1$, are different in the intermediate region.

Let us show that the stationary state defined by (9.13) and (9.14) is stable. Consider the disturbance

$$K=K_0+\delta K\,t^\lambda,\quad Q=Q_0+\delta Q\,t^\lambda,$$

$$a=\frac{\alpha_0+\delta\alpha\,t^\lambda}{t},\quad b=\frac{\beta_0+\delta\beta\,t^\lambda}{t}.$$

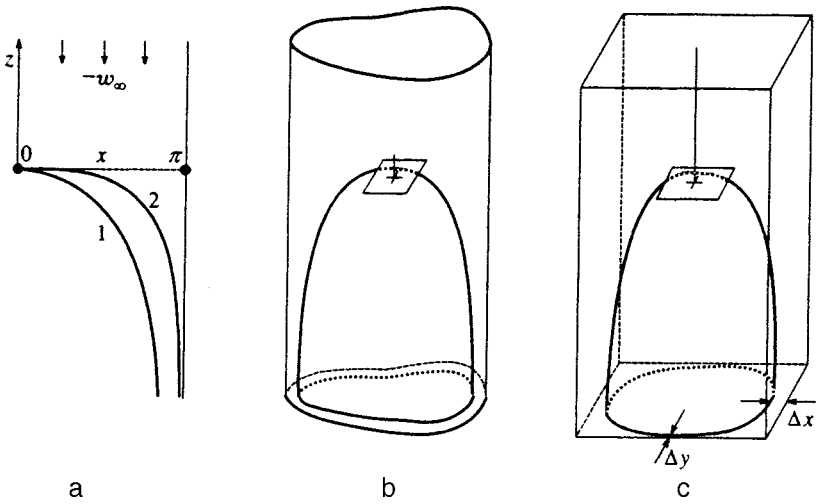

FIG. 15. One-parameter families of bubbles. a—2D case, b,c—3D case.

We linearize the system (9.7)–(9.10) near it. The matrix M for the eigenvalues λ has the form

$$\begin{matrix} -\lambda-3\alpha-q\beta & 0 & 1-3K & -qK \\ 0 & -\lambda-\alpha-3q\beta & -Q & q^2-3qQ \\ \alpha+\beta & 0 & (1-K)(\lambda-1)+2\alpha & K(1-\lambda) \\ 0 & \alpha+\beta & Q(1-\lambda) & (q-Q)(\lambda-1)+2q^2\beta. \end{matrix}$$

Here we have omitted the subscript 0 in the stationary functions for brevity. It follows from the calculations of the eigenvalues λ that in the interval occupied by curve L with the end points T and b $(1<q<1.26)$ all the eigenvalues are different and negative.[11)] Therefore, there is power-law damping of the disturbances and the stationary state defined by (9.13) and (9.14) is stable. We note that on branch Db, which is depicted by dashed lines in Figs. 14(a) and 14(c), there is one positive eigenvalue λ, which is an indication of the instability of this stationary state.

As in the case of Rayleigh–Taylor instability, the trajectories of the system (9.7)–(9.10) for $g=0$ connect the vicinity of the center at $K=Q=0$ and the node S. Qualitatively, the structure of the phase space is similar to that shown in Fig. 13. In particular, for $q=1$ and asymmetric initial data $[a(0)\neq b(0)]$ there is an increase in symmetry $(2m\rightarrow 4m)$ as $t\rightarrow\infty$. The two-dimensional phase space of the square lattice (Secs. 4.5 and 4.6) is located on the bisector $K=Q$, $a=b$ of the four-dimensional phase space of the system (9.7)–(9.10).

10. ONE-PARAMETER CHARACTER OF RAYLEIGH–TAYLOR STATIONARY STATES

The nonstationary stage was investigated above. It was shown that during evolution, the growth of disturbances is saturated and a weakly perturbed state transforms into a stationary (limiting) state. Because of this, the theory of stationary states is important. Let us examine it in the case of Rayleigh–Taylor instability. The principal property is the one-parameter $(1d)$ character of the stationary states.[41–43,24] Owing to it, a quantitative theory of stationary states has been devised.[24]

10.1. One-parameter solutions in the 2D geometry. Let us consider the potential streamlining by an external flow with the profile $z=\eta(x)$ depicted in Fig. 15(a) (curve *1*). The profile is found in the band $0<x<\pi$. The freestream velocity $w_\infty=1$ is assigned for $z\rightarrow+\infty$. The function $\eta(x)$ decreases monotonically, and $\eta\rightarrow-\infty$ as $x\rightarrow\pi$. The form of the profile $\eta(x)$ assigns the streaming potential $f(\xi=x+iz)$. The pressure distribution on the profile boundary $p_\eta(x)=p[x,z=\eta(x)]$ is determined together with the potential. The position of a point on this boundary is assigned both by the x coordinate and by the z coordinate. Therefore, the distribution of the pressure can be represented in the form of the function $p_\eta(z)$. Thus, the profile η specifies the distribution $p_\eta(z)$ $(\eta\rightarrow p_\eta)$.

In the opposite case the function $p_\eta(z)$ is assigned (as before, $w_\infty=1$). There is a mutual one-to-one correspondence between $p_\eta(z)$ and $\eta(x)$ (at least locally). Therefore, assigning $p_\eta(z)$ fixes the potential $f(\xi)$ and the form of $\eta(x)$ $(p_\eta\rightarrow\eta)$.

Let us examine the dynamics of a heavy fluid with a free surface in a uniform gravitational field. In this context we are interested only in the special class of the distributions $p_\eta(z)$ consisting of the linear distributions

$$p_\eta(z)=gz \tag{10.1}$$

with $g>0$. There is a single parameter g, which runs through the class (10.1). The distributions (10.1) correspond to a one-parameter class of the potentials $f(\xi;g)$ and the profiles $\eta(x;g)$. They are also solutions of the problem of the stationary rise of a bubble with the boundary $\eta(x;g)$, since this boundary is a streamline and the condition $(u^2+w^2)|=-2g\eta$ holds on it by virtue of (10.1) and Bernoulli's equation.

Therefore, the problem of stationary Rayleigh–Taylor bubbles has a $1d$ family of solutions. Figure 15(a) shows two bubbles from this family (curves *1* and *2*), which correspond to the values g_1 and $g_2>g_1$.

It is clear that the velocity w_∞ can serve as the parameter instead of g. In fact, let us fix the form of a certain profile $\eta(x;g_f)$ from the class (10.1). When the form of η is fixed, the variation of g (the pressure is measured relative to the pressure at the apex of the profile) leads to the variation of

w_∞. A more obtuse bubble (curve *2* in [Fig. 15(a)] has a smaller rise velocity w_∞.

The existence of a one-parameter family of solutions in the heavy-fluid problem is well known from the theory of gravity waves.[44]

10.2. One-parameter solutions in the 3D geometry. It follows from simple arguments that there are also one-parameter solutions in the 3D case. Let us demonstrate this. Consider the potential streamlining of the profile $\eta(x,y)$ in a vertical pipe of arbitrary cross section [Figs. 15(b) and 15(c)]. Let $w_\infty=1$. The potential $f(x,y,z)$ and the pressure distribution $p_\eta(x,y)$ on the boundary are specified by the form of the profile $\eta(x,y)$ ($\eta\to p_\eta$).

We represent the function $p_\eta(x,y)$ in the form of the equivalent function $p_\eta(z,\theta)$, where θ is the azimuthal angle lying in the horizontal plane. The profiles η have a stagnation point. It is marked by a cross in Figs. 15(b) and 15(c). We shall call it the bubble apex B. The case in which the point B is the highest point on the profile has bearing on the heavy-fluid problem. Accordingly, the tangent plane at that point is horizontal. A streamline entering an apex is perpendicular to the tangent plane. The crosses in Figs. 15(b) and 15(c) indicate the directions of the principal curvatures K and Q.

In the opposite case the function $p_\eta(z,\theta)$ is assigned. It specifies the potential $f(x,y,z)$ and the form of $\eta(x,y)$.

A uniform gravitational field corresponds to a special class of the functions $p_\eta(z,\theta)$. In it these functions are isotropic and linear:

$$p_\eta(z,\theta)=gz. \tag{10.2}$$

This is a simple consequence of Bernoulli's theorem, which must be applied to the streamlines emanating from an apex and to the contours which are obtained in horizontal sections of the surface η. The class (10.2) is associated with a $1d$ family of potentials and bubbles, which can be represented in the form $f(x,y,z;g),\eta(x,y;g)$ and in the form $f(x,y,z;w_\infty),\eta(x,y;w_\infty)$.

Let w_∞ be the parameter. As in the 2D case [compare curves *1* and *2* in Fig. 15(a)], a bubble with a small value of w_∞ is more obtuse (the curvatures K and Q are smaller) and has a thinner near-wall jet [Figs. 15(b) and 15(c)].[12)]

The variety of pipe cross sections is greater in the 3D geometry. In the 2D case the pipe is a band or a slot bounded by two parallel walls, which are perpendicular to the x,z plane [see Fig. 15(a)]. In the 2D case the bubbles have the form of shafts perpendicular to the plane of the figure, and in the 3D case they have the form of ''fingers'' surrounded by near-wall jets. The special cases of pipes of general form correspond to pipes with symmetric cross sections, which can serve as the basis for parquet covering [Sec. 3, Figs. 1–6 and 15(c)]. In these cases the potentials are given by the series (3.1) and (3.2), and the corners and walls are symmetric verticals and planes (Sec. 3).

11. TRUE STATIONARY STATES AND ONE-PARAMETER SOLUTIONS

The parabolic approximation ($N=1$, Sec. 1) is fairly simple (Secs. 4 and 9). For this reason, it is possible to analytically describe structural effects associated with the 3D geometry (Secs. 5–7). The extension of this approach to higher approximations ($N>1$) is a very interesting problem. When $N=1$, the system undergoes a transition of an initial, weakly disturbed state I to a stationary state S ($I\to S$; Secs. 4, 5, and 9). The trajectory $I\to S$ eludes the surface G (Fig. 13). The analysis of systems of higher order that are similar to (4.17) runs into difficulties. It turns out for $N>1$ that trajectories which start out from I states are blocked by the surface G.[13,15] We can approach the problem from another side. Consider the stationary states and the trajectories near them. In this way it is possible to obtain some important results.

Zero-parameter solutions. An analysis of the vicinity of stationary states requires finding stationary points and investigating their stability. Let us consider some aspects of the search for stationary states. From the qualitative standpoint the situation in the 2D and 3D cases is completely identical.[28,29,17] In higher approximations in the 2D geometry we have

$$\eta=\sum_{n=1}^{N} K_n x^{2n},\quad f(\xi)=\varphi+i\psi=-\sum_{n=1}^{N} a_n\left(\frac{e^{in\xi}}{n}-i\xi\right),$$
$$\xi=x+iz. \tag{11.1}$$

Stationary solutions are obtained from the algebraic systems

$$\psi_\alpha(K_1,\ldots,K_N,a_1,\ldots,a_N)=0,$$
$$p_\alpha(K_1,\ldots,K_N,a_1,\ldots,a_N)=0, \tag{11.2}$$

where $\alpha=1,\ldots,N$. When $N=1$, the system (11.2) for unknown K_n and a_n (11.1) takes the form $\psi_1=1-3K_1=0$, $p_1=a_1^2-K_1=0$. It is similar to the equations $\psi_1=1-4K_1=0$, $p_1=a_1^2-4K_1=0$, which are obtained from the system (4.17) ($g=1;6m,4m$) when $\dot K_1=0$ and $\dot a_1=0$.

When $N\leqslant 6$, the systems (11.2) can be solved exactly.[45] Let the index i run through the roots of the system (11.2) $(K_N,a_N)^i$ in the Nth approximation, where $K_N=\{K_1,\ldots,K_N\}$ and $a_N=\{a_1,\ldots,a_N\}$. It was shown in Ref. 17 that from the set of roots $\{i\}$, no more than one root is physically correct in each order defined by N in the range $N\leqslant 6$. We denote this root by

$$(K_N,a_N)^{0d}. \tag{11.3}$$

A program for numerically solving the system (11.2) by Newtonian iterations was written. It was found that the distances between the roots (11.3) from different orders as defined by N are significantly greater than the convergence radius of the iterations. It was also found that iterations do not permit reaching an approximation order N with a value $N>6$ when roots with $N\leqslant 6$ are used as the initial approximation.

One-parameter solutions. An interesting approach to the problem of stationary states is associated with their one-parameter character. Section 10 was devoted to a qualitative

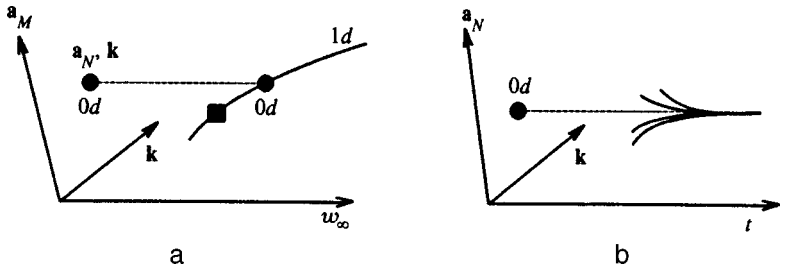

FIG. 16. a) Curve $1d$ and point $0d$ on it. b) The evolution of the system at $t\to\infty$ can end only at point $0d$, solid curves–trajectories of the system (11.10).

proof of the existence of one-parameter solutions. Let us analyze it quantitatively. The one-parameter solutions are described by the expansions

$$\eta=\sum_{n=1}^{N} K_n x^{2n}, \quad f(\xi)=-\sum_{m=1}^{M} a_m \frac{e^{im\xi}}{m}+iw_\infty\xi, \quad (11.4)$$

where N is the approximation order of the boundary conditions. The spectral decomposition (11.4) has the asymptote needed for $z\to+\infty$. The expression

$$M=N+1, \quad (11.5)$$

which relates the number of harmonics in the $1d$ solution (11.4) and the approximation order, is important. The system of $2N+1$ equations which are satisfied by the unknown $K_1,\ldots,K_N,a_1,\ldots,a_N,a_M$, has the form

$$\begin{aligned}&\psi_\alpha(K_1,\ldots,K_N,a_1,\ldots,a_N,a_M)=0,\\ &p_\alpha(K_1,\ldots,K_N,a_1,\ldots,a_N,a_M)=0,\\ &a_1+\ldots+a_N+a_M=w_\infty, \quad \alpha=1,\ldots,N,\end{aligned} \quad (11.6)$$

where the number of the last Fourier amplitude taken into account in a particular order N is given by formula (11.5) and the parameter w_∞ has been eliminated from the functions ψ_α and p_α using the last equation in the system (11.6). The condition for the presence of the required stagnation point at the apex $\xi=0$ supplements the equation $\Sigma a_m=w_\infty$ in the system (11.6). The solutions of Eqs. (11.6), as they should be (Sec. 10), are the functions

$$K_1(w_\infty),\ldots,K_N(w_\infty),a_1(w_\infty),\ldots,a_N(w_\infty),a_M(w_\infty) \quad (11.7)$$

of the parameter. We shall call the curves (11.7) one-parameter or $1d$ solutions, and we shall call a solution in the form of the point (11.3) a zero-parameter or $0d$ solution.

A program was written for the iterative solution of the system (11.6). In contrast to the case of (11.2), it was found that the iterations converge rapidly. This opens up a way to find $0d$ points using $1d$ curves.

Relation between $0d$ *solutions and* $1d$ *solutions*. Compare the systems (11.2) and (11.6) along with (11.5) and their solutions (11.3) and (11.7) for identical values of N. As stated, the $1d$ solutions (11.7) are functions of w_∞. In particular, the last amplitude a_M is a function of w_∞. It turns out that there is a value of the parameter w_∞^{0d} at which

$$a_M(w_\infty^{0d})=0. \quad (11.8)$$

Let

$$K_1^{0d},\ldots,K_N^{0d},a_1^{0d},\ldots,a_N^{0d},a_M^{0d}=0 \quad (11.9)$$

be the solution (11.7) for w_∞^{0d}. We see that it belongs to curve (11.7).

Let us compare Eqs. (11.2) and (11.6). We substitute the solution (11.9) without a_M^{0d} into the system (11.2) of the same order. It is easy to see that (11.9) gives a solution of the system (11.2) of the form (11.3).

This allows us to circumvent the difficulty associated with the absence of convergence of the iterative solution of (11.2), since the system (11.2) and the system (11.6) without the last amplitude a_M and without the last equation $\Sigma a_m=w_\infty$ are identical. More specifically, the $1d$ curve of the solutions (11.7) or the system (11.6) is constructed iteratively. The point (11.8) is sought on it. This point is the $0d$ point in the respective order as defined by N. The following $0d$ points were found by this method: $R(N=1)=3$, $R(N=3)=2.57$, $R(N=6)=2.556$, and $R(N=8)=2.4445$, where R is the radius of curvature at the bubble apex.

Figure 16(a) shows the $1d$ curve (11.7) and the $0d$ point (11.3), (11.8), (11.9) lying on it for a certain order N. We introduce the notation $\mathbf{K}=K_1,\ldots,K_N$, $\mathbf{a}_M=a_1,\ldots,a_N,a_M$, and $\mathbf{a}_N=a_1,\ldots,a_N$, and the Mth component of the vector $\mathbf{a}_N$ is equal to zero. The projection of the $0d$ point onto the $\mathbf{K}$, $\mathbf{a}_M$ space lies in the $a_M=0$ plane. The projections of all the other points from the $1d$ curve onto this space, for example, the point marked by a square in Fig. 16(a), have $a_M\neq0$.

True stationary states. It is not difficult to understand that only the $0d$ points are true stationary states. In fact, a nonstationary system of the type (4.17) takes the following form in higher orders:

$$\begin{aligned}&\dot K_\alpha=\psi_\alpha(\mathbf{K},\mathbf{a}_N), \quad \sum_{\beta=1}^{N} G_{\alpha\beta}\dot a_\beta=p_\alpha(\mathbf{K},\mathbf{a}_N),\\ &\alpha=1,\ldots,N,\end{aligned} \quad (11.10)$$

where the right-hand sides ψ_α and p_α are the same as in the system (11.2).[13,15] The system (11.2) is obtained from Eqs. (11.10) when $\dot K=0$ and $\dot a=0$. Therefore the solution (11.3) is a stationary point of the system (11.10). The solutions of Eqs. (11.10) asymptotically approach the limiting values (11.3), (11.8), and (11.9) [see Fig. 16(b)]. In addition, in the asymptote the sum $\Sigma^N a_n$ takes the value (w_∞^{0d}) which is obtained in that limit, i.e., it cannot be required that this value differ from w_∞^{0d}. The stability of the $0d$ point has been demonstrated for $N=1$ (see Secs. 4 and 9 and Refs. 15 and 18). The $0d$ points are probably stable when $N>1$.

We reiterate that the remaining points from the $1d$ curve, for example, the square in Fig. 16(a), are not station-

ary points in the respective order as defined by N. They cannot be investigated for stability. If a $1d$ solution (11.7), say the square, is taken, the last amplitude a_M is discarded, and the ordered set $K_1,\ldots,K_N,a_1,\ldots,a_N$ thus obtained is substituted into the nonstationary system (11.10), the right-hand sides of these equations will not be equal to zero, and, therefore, there will be a deviation from stationary behavior ($\dot{K}\neq 0,\dot{a}\neq 0$) at that point.

In this respect, stationary bubbles [$g=1$, $\partial_t=0$, and boundary conditions (2.1) and (2.2)] differ qualitatively from stationary gravity waves ($g=-1$, $\partial_t=0$, and the same boundary conditions). In the case of waves we have

$$\eta=\sum_{n=1}^{N} K_n x^{2n}, \quad f(\xi)=i\sum_{n=1}^{N} a_n \frac{e^{in\xi}}{n}-w_\infty\xi,$$

the amplitudes a_n are real, and there is fluid at $z>\eta$. The system for waves which is analogous to system (11.6) has the form

$$\hat{\psi}_\alpha(K_1,\ldots,K_N,a_1,\ldots,a_N;w_\infty)=0,$$
$$\hat{p}_\alpha(\ldots)=0, \quad \alpha=1,\ldots,N, \tag{11.11}$$

where the carets above ψ and p indicate that the respective functions of the arguments differ from the functions in (11.6). It is significant that there is no additional equation associated with the stagnation point here, and, therefore, there is no additional amplitude a_M. The solutions of the system (11.11) are functions of the parameter:[44]

$$\hat{K}_1(w_\infty),\ldots,\hat{K}_N(w_\infty), \quad \hat{a}_1(w_\infty),\ldots,\hat{a}_N(w_\infty). \tag{11.12}$$

Because the number of nonstationary equations in the wave case that are analogous to Eqs. (11.10) coincides with the number of the functions (11.12), linearization of the nonstationary equations can be performed near each point on the $1d$ curve (11.12).

Uniqueness of the $0d$ *stationary point.* It is important to stress that in each order as defined by N there is no more than one $0d$ point. Therefore, the solution of the asymptote problem for Rayleigh–Taylor bubbles is unique.

We express our sincerest thanks to S.I. Anisimov and O.M. Belotserkovskiĭ for their interest in this work. This study was carried out with financial support from the Russian Fund for Fundamental Research (Projects 99-02-16666 and 97-01-00931) and the program for supporting leading scientific teams (96-15-96448 and 96-15-96137).

*)E-mail: nail@landau.ac.ru
†)E-mail: oparin@landau.ac.ru

1)The experimental and numerical studies are cited over the course of the presentation.

2)The linear stage is symmetric. In it an arbitrary disturbance is a linear combination of harmonics. The harmonics develop independently. In each sinusoidal period the bubbles and jets are symmetric.

3)A preliminary description of these effects was given in Ref. 18. Only the $p4mm$ case (a square) was considered in that work. The general case ($pmm2$, $p6mm$, and $p3m1$ lattices; for the symmetry classes see Refs. 25–27) is considered below. We note the following new effects: 1) the identical nature of the bubble dynamics in different lattices; 2) an asymptotic increase in symmetry (the symmetry of the set $\{k\}$ of the initial wave vectors is more important than the collective symmetry of the set of initial amplitudes and vectors $\{a;k\}$); 3) interchangeability of hexagonal and triangular bubbles upon rephasing of Richtmyer–Meshkov and Rayleigh–Taylor instability. Incidentally, the array of crests is hexagonal during half of a period according to the time of a standing gravity wave and is triangular during the other half.

4)This also permits the use of parabolic model 1 to describe flow.

5)The principal curvatures have the opposite sign at saddle points.

6)In the case of a rectangular bubble the principal curvatures at the B and J apices differ in value, and the contour of a cross section of the surface η in a horizontal plane is an ellipse (see Fig. 1).

7)The stationary point (4.18), which applies to Rayleigh–Taylor instability, was previously calculated in Refs. 8, 17, 18, 28, and 29 for the case of a square array and in Ref. 8 for the case of a hexagonal array.

8)In it the smallest possible number of symmetric walls (three) converge at a jet apex J. Just this feature sets apart the $6m$ bubble array. The bubbles in a $6m$ lattice differ little from the bubbles in a $4m$ lattice.

9)This approach can also be applied to a J apex and to S_s and S_w saddle points (see Sec. 5.3). At these points, just as at a B point, the linear terms drop out, and the expansion of the boundary η begins with the quadratic terms.

10)We note that in the case of high-symmetry lattices (Sec. 4) there is no need to investigate the stability of the Rayleigh–Taylor and Richtmyer–Meshkov stationary states. It is obvious from the structure of the phase space. Conversely, in the $2m$ case an investigation of the stability yields a description of this structure.

11)For example, when $q=1$, we have $\det M=2\lambda^4+10\lambda^3+17\lambda^2+11\lambda+2=0$ with the eigenvalues -1, -2, and $(-4\pm 2\sqrt{2})/4$. The linearization of (4.17) gives the system $(\lambda+2)\delta K=0$, $2\delta K+(\lambda+1)\delta a=0$ with the roots $\lambda_1=-1$ and $\lambda_2=-2$. When $q\neq 1$, the characteristic equation $\det M=0$ is extremely cumbersome.

12)We recall that the limiting case ($t\gg 1/\sqrt{gk}$), in which the S_s and S_w saddle points and J apices are at large distances from the B apex, is being considered. The ratio Δ_x/Δ_y between the thicknesses of the near-wall jets is discussed in Sec. 5.3.

[1] A. M. Prokhorov, S. I. Anisimov, and P. P. Pashinin, Usp. Fiz. Nauk **119**, 401 (1976) [Sov. Phys. Usp. **19**, 547 (1976)].
[2] S. I. Anisimov, A. M. Prokhorov, and V. E. Fortov, Usp. Fiz. Nauk **142**, 395 (1984) [Sov. Phys. Usp. **27**, 181 (1984)].
[3] M. M. Marinak, S. G. Glendinning, R. J. Wallace, B. A. Remington *et al.*, Phys. Rev. Lett. **80**, 4426 (1998).
[4] M. Herant and S. E. Woosley, Astrophys. J. **425**, 814 (1994).
[5] W. Hillebrandt, Astrophys. J. **452**, 769, 779 (1995).
[6] S. I. Anisimov, Ya. B. Zel'dovich, N. A. Inogamov, and M. F. Ivanov, in *Shock Waves, Explosions and Detonation (Progress in Astronautics and Aeronautics Series, Vol. 87)*, J. R. Bowen, J.-C. Leyer, and R. I. Soloukhin (Eds.), AIAA, Washington, D. C. (1983), p. 218.
[7] K. O. Mikaelian, Phys. Rev. Lett. **80**, 508 (1998).
[8] S. I. Abarzhi, Phys. Rev. Lett. **81**, 337 (1998).
[9] Q. Zhang, Phys. Rev. Lett. **81**, 3391 (1998).
[10] V. N. Goncharov, Phys. Rev. Lett. **82**, 2091 (1999).
[11] D. Layzer, Astrophys. J. **122**, 1 (1955).
[12] R. M. Davies and G. I. Taylor, Proc. R. Soc. London, Ser. A **200**, 375 (1950).
[13] N. A. Inogamov, Pis'ma Astron. Zh. **20**, 754 (1994) [Astron. Lett. **20**, 651 (1994)].
[14] J. Hecht, U. Alon, and D. Shvarts, Phys. Fluids **6**, 4019 (1994).
[15] N. A. Inogamov, Zh. Éksp. Teor. Fiz. **107**, 1596 (1995) [JETP **80**, 890 (1995)].
[16] U. Alon, J. Hecht, D. Ofer, and D. Shvarts, Phys. Rev. Lett. **74**, 534 (1995).
[17] N. A. Inogamov, Laser Part. Beams **15**, 53 (1997).
[18] N. A. Inogamov, in *Proceedings of the 6th International Workshop on the Physics of Compressible Turbulent Mixing*, G. Jourdan and L. Houas (Eds.), Institut Universitaire des Systèmes Thermiques Industriels, Marseille (1997), p. 208.
[19] S. W. Haan, Phys. Fluids B **3**, 2349 (1991).

[20] A. L. Velikovich and G. Dimonte, Phys. Rev. Lett. **76**, 3112 (1996).
[21] R. Menikoff and C. Zemach, J. Comp. Physiol. **51**, 28 (1977).
[22] G. R. Baker and D. W. Moore, Phys. Fluids A **1**, 1451 (1989).
[23] A. I. Dyachenko, E. A. Kuznetsov, M. D. Spector, and V. E. Zakharov, Phys. Lett. A **221**, 73 (1996).
[24] N. A. Inogamov, JETP Lett. **55**, 521 (1992).
[25] A. V. Shubnikov, *Symmetry* [in Russian], Izd. Akad. Nauk SSSR, Moscow–Leningrad (1940).
[26] A. V. Shubnikov and V. A. Koptsik, *Symmetry in Science and Art*, Plenum Press, New York (1974) [Russ. original, Nauka, Moscow (1972)].
[27] Yu. I. Sirotin and M. P. Shaskol'skaya, *Fundamentals of Crystal Physics*, Mir, Moscow (1982) [Russ. original, Miṛ, Moscow (1979)].
[28] S. I. Abarzhi and N. A. Inogamov, Zh. Eksp. Teor. Fiz. **107**, 245 (1995) [JETP **80**, 132 (1995)].
[29] N. A. Inogamov and S. I. Abarzhi, Physica D **87**, 339 (1995).
[30] Yu. M. Davydov and M. S. Panteleev, Prikl. Mekh. Tekh. Fiz. **1**(125), 117 (1981).
[31] D. L. Youngs, Phys. Fluids A **3**, 1312 (1991).
[32] X. L. Li, Phys. Fluids **8**, 336 (1996).
[33] J. F. Haas, I. Galametz, L. Houas, G. Jourdan, and G. Rodriguez, Chocs, No. **14**, 15 (1995).
[34] A. N. Aleshin, S. G. Zaĭtsev, E. V. Lazareva, V. B. Rozanov, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza **6**, 111 (1995).
[35] M. D. Kamchibekov, E. E. Meshkov, N. V. Nevmerzhitsky, and E. A. Sotskov, in *Proceedings of the 6th International Workshop on the Physics of Compressible Turbulent Mixing*, G. Jourdan and L. Houas (Eds.), Institut Universitaire des Systèmes Thermiques Industriels, Marseille (1997), p. 238.
[36] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed., Pergamon Press, Oxford (1987) [Russ. original, Nauka, Moscow (1986)].
[37] O. M. Belotserkovskiĭ, *Numerical Simulation in the Mechanics of Continuous Media* [in Russian], Nauka, Moscow (1984).
[38] O. M. Belotserkovskiĭ, *Numerical Experiments in Turbulence: from Order to Chaos* [in Russian], Nauka, Moscow (1994).
[39] O. M. Belotserkovskiĭ, V. A. Gushchin, and V. N. Kon'shin, Zh. Vychisl. Mat. Mat. Fiz. **27**, 594 (1987).
[40] S. I. Anisimov, A. V. Chekhlov, A. Yu. Dem'yanov, and N. A. Inogamov, Russ. J. of Comput. Mech. **1**(2), 5 (1993).
[41] P. R. Garabedian, Proc. R. Soc. London, Ser. A **241**(1226), 423 (1957).
[42] G. Birkhoff and D. Carter, J. Math. Mech. **6**, 769 (1957).
[43] J.-M. Vanden-Broeck, Phys. Fluids **27**, 1090 (1984).
[44] L. W. Schwartz and J. D. Fenton, Annu. Rev. Fluid Mech. **14**, 39 (1982).
[45] N. A. Inogamov and A. V. Chekhlov, Dokl. Ross. Akad. Nauk **328**, 311 (1993) [Dokl. Phys. **38**, 32 (1993)].

Translated by P. Shelnitz

Diffraction of x rays at a Bragg angle of $\pi/2$ (back reflection) with consideration of multiwave effects

V. G. Kohn,*) I. V. Kohn, and É. A. Manykin

Russian Research Center ''Kurchatov Institute'', 123182 Moscow, Russia
(Submitted 26 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 940–952 (September 1999)

The energy dependence of the back reflectivity in the dynamical diffraction of x rays at a Bragg angle of $\pi/2$ (back diffraction) in perfect crystals of cubic symmetry (silicon) is investigated theoretically. In this case strict backscattering is realized only under the conditions of multiple diffraction. The features of the influence of multiple diffraction on back reflection in the energy range near the nuclear resonance radiation energy of 14.41 keV for ^{57}Fe nuclei, specifically in the six-wave case, including the silicon (1,9,9) reflection (with an energy of 14.57 keV), which can be investigated experimentally with high energy resolution (1 meV) using synchrotron radiation and a monochromator developed for nuclear resonant absorption, are thoroughly studied. It is shown that the back reflectivity observed under the conditions of multiple diffraction has several maxima on the plot of its energy dependence with a value at each maximum smaller than half, in contrast to two-wave diffraction, where there is one maximum with a value close to unity. © *1999 American Institute of Physics.* [S1063-7761(99)01309-8]

1. INTRODUCTION

The back reflection of x rays during diffraction on perfect crystals with a Bragg angle of $\pi/2$ (back diffraction) is known to occur only in a very narrow energy range with a relative width less than 10^{-6}, but has a relatively weak sensitivity to the angular divergence of the beam (no more than 10^{-3} rad). Since the construction of the dispersion surface introduced into the theory by Ewald[1] is impossible in the case under consideration, it initially appeared that a generalized solution of Maxwell's equations without linearization of the dispersion correction to the wave vector must be used to analyze back reflection.[2,3] In reality, the theory remains linearized to a high accuracy, and in terms of the deviation parameter from the Bragg condition it does not differ in any way from diffraction at a Bragg angle smaller than $\pi/2$ (Ref. 4).

A slight angular deviation of the beam for the direction corresponding to strict backscattering was used in the experimental investigations of back reflection in Refs. 5–7, since otherwise it was impossible to pass the incident beam through the opaque detector. This simultaneously permitted elimination of the multiwave effects and allowed the use of the theory of two-wave diffraction to describe the measured plots of the energy (temperature) and angular dependences of the reflectivity. Additional back reflection was employed to monochromatize the beam, and the convolution of two theoretical reflection curves was calculated simultaneously for comparison with experiment. Good agreement between the experimental temperature curve and the theoretical calculation was obtained in Ref. 7.

Nevertheless, strict backscattering (back diffraction) is of considerable interest in connection with the possibility of using it to create an x-ray analog of the familiar Fabry–Pérot interferometer (see, for example, Ref. 8 and the references therein). In this case very high monochromatization of the radiation is needed to ensure a long longitudinal (temporal) coherence length. The necessary degree of monochromatization is achieved with a safety margin using a ''Mössbauer monochromator,'' i.e., the nuclear resonant scattering of pulses of synchrotron radiation in conjunction with a time-window technique, under which a detector with a high temporal resolution, of the order of a nanosecond, permits isolation of only the scattered radiation delayed by nuclei. The latter has an energy width of the order of the width of the excited state of the nuclei Γ. Only the ^{57}Fe nuclear transition with an energy $E=14.413\,\text{keV}$ and a width $\Gamma=4.66\times10^{-6}\,\text{meV}$ has been used hitherto fairly widely. A transparent detector and a large crystal–detector distance must also be employed to measure strict backscattering. The incident (primary) synchrotron radiation pulse is also cut off using a time window.

Just such a measurement technique was recently first proposed and successfully implemented in Ref. 9. Sapphire (Al_2O_3) crystals, which did not have a sufficiently perfect crystal lattice, were used to eliminate the multiwave effects in Ref. 9. For this reason, despite the high angular collimation and the very high monochromaticity of the incident beam, the experimental curves differed from the results of a calculation based on the dynamical theory for perfect crystals.

Hitherto, only silicon crystals had a sufficiently perfect structure. In this case several reflections have energies close to $E=14.413\,\text{keV}$. They are the (3,5,11) reflection with $E=14.210\,\text{keV}$, the (0,4,12) reflection with $E=14.437\,\text{keV}$, and the (1,9,9) reflection with $E=14.572\,\text{keV}$. In all cases back reflection is realized under the conditions of multiple diffraction. The reflections indicated were recently measured in Ref. 10 at room tempera-

ture. An x-ray monochromator with a resolution of the order of 1 meV was used this time. Monochromators of such a type were widely used in the last few years in the Mössbauer facilities of third-generation synchrotron radiation sources (ESRF in France, APS in the U.S.A., and SPring-8 in Japan), in inelastic nuclear resonant absorption experiments (for the latest results on this subject, see Refs. 11 and 12 and the references therein) and were developed specifically for $E = 14.4\,\text{keV}$ with the possibility of scanning the energy in a small range.

Moreover, the use of back reflectivity peaks in silicon as reference marks on the energy scale of such a monochromator permits measurement of the energy of the nuclear transition itself to a higher accuracy in comparison to other methods. Just such a problem was solved in Ref. 10. For this purpose, in particular, it is necessary to know how the multiwave effects influence the form of the back reflectivity peak. Thus, an investigation of strict backscattering with consideration of the multiwave effects has practical value in addition to being of purely physical interest. The (3,5,11) and (0,4,12) reflections correspond to 24-wave diffraction. They will be studied at a later date. The present work is devoted to an analysis of back reflection with consideration of multiwave effects in the case of the silicon (1,9,9) reflection, which corresponds to 6-wave diffraction. The dynamical theory of the diffraction of plane waves in matrix form and the scheme for the computer calculations are presented in the next section. The scattering geometry and the results of the numerical calculations are presented in Sec. 3. Section 4 offers a qualitative analysis of the influence of multiwave corrections on two-wave diffraction in ranges of parameters where they can be regarded as a perturbation. It provides partial explanations for the numerical results obtained.

2. MATRIX FORM OF THE DYNAMICAL THEORY OF THE MULTIPLE DIFFRACTION OF PLANE WAVES

The theory is devised for a monochromatic plane wave with a frequency ω and a wave vector $\mathbf{K}_0$. Real radiation can always be represented as a superposition of plane waves, and we assume that the different frequencies and directions of the wave vectors are incoherent. Thus, the intensity of the back-reflected radiation for a monochromatic plane wave must be calculated, and then the result must be averaged over the finite angular and frequency (energy) ranges corresponding to the results of the specific experiment. When the conditions for multiple diffraction in a crystal in the form of a plane-parallel plate with an internal normal $\mathbf{n}$ to the entrance surface of the crystal are satisfied, an incident wave with the electric field intensity

$$\mathbf{E}_0(\mathbf{r},t)=\mathbf{E}_0\exp(i\mathbf{K}_0\cdot\mathbf{r}-i\omega t) \tag{1}$$

corresponds to the superposition of truncated Bloch waves:

$$\mathbf{E}(\mathbf{r},t)=\sum_j \lambda_j \sum_m \mathbf{E}_{mj}\exp(i\mathbf{k}_{mj}\cdot\mathbf{r}-i\omega t),$$

$$\mathbf{k}_{mj}=\mathbf{K}_0+\mathbf{h}_m+\varepsilon_j\mathbf{n}, \tag{2}$$

which contains only reciprocal-lattice vectors $\mathbf{h}_m$ of the crystal that satisfy the Bragg condition $(\mathbf{K}_0+\mathbf{h}_m)^2\approx\mathbf{K}_0^2$ to an assigned accuracy of the order of the amplitude χ_m of the diffraction scattering from one wave to another. The subscript j labels the possible solution, and λ_j is the degree of excitation of the respective solution in the crystal for an assigned incident wave. It is found from the boundary conditions.

When solutions in the form (2) are plugged into Maxwell's equation for the amplitude of the electric field, the following approximations are made to an accuracy of the order of $\chi_0\approx10^{-6}$.

1) The electric field is assumed to be transverse:

$$\mathbf{E}_{mj}=\sum_s E_{msj}\mathbf{e}_{ms}, \tag{3}$$

where $s=\pi,\sigma$ is the polarization index, and the unit vectors $\mathbf{e}_{ms}$ specify the polarization direction in beam m in a plane perpendicular to the unit vector $\mathbf{s}_m$, which is parallel to $\mathbf{K}_0+\mathbf{h}_m$.

2) Only the first power of the dispersion correction ε is taken into account in the equations. This corresponds to the approximation of generalized geometric optics in the small-angle case.

3) Averaging of the equation over a unit cell of the crystal is performed for the purpose of eliminating the fast variables with a variation length of the order of the wavelength of x rays from the calculations.

4) Only the dipolar interaction of the electromagnetic wave with the medium is taken into account (the accuracy of this approximation is poorer than that of the preceding approximations, but in all cases, except the anomalous transmission effect, it is sufficient).

The approximations indicated allow us to write equations separately for each of the scalar amplitudes E_{msj} in the following form (for further details, see Refs. 1 and 13):

$$\left(\frac{\gamma_m}{K}\varepsilon+\alpha_m\right)E_{ms}=\sum_{m',s'} g_{mm'}^{ss'}E_{m's'}, \tag{4}$$

where $K=\omega/c$ is the wave number, c is the speed of light,

$$\gamma_m=(\mathbf{s}_m\cdot\mathbf{n}),\quad g_{mm'}^{ss'}=\chi_{m-m'}(\mathbf{e}_{ms}\cdot\mathbf{e}_{m's'}),$$

$$\alpha_m=[(\mathbf{K}_0+\mathbf{h}_m)^2-\mathbf{K}_0^2]/K^2, \tag{5}$$

and $\chi_{m-m'}$ is the Fourier component of the polarizability of the crystal in the reciprocal-lattice vector $\mathbf{h}_m-\mathbf{h}_{m'}$.

To describe the calculation scheme in matrix form it is convenient to combine the two indices m and s into one, for which we retain the notation m. The index m thus runs through the values $0\pi,0\sigma,1\pi,1\sigma,\ldots,(n-1)\pi,(n-1)\sigma$ in the n-wave case. Going over to the new amplitudes, $B_{mj}=\gamma_m^{1/2}E_{mj}$, we can rewrite the system of equations (4) in the form characteristic of many dynamical systems (electrons, phonons, etc.), i.e., as the eigenvalue problem

$$\varepsilon B_m=\sum_{m'} G_{mm'}B_{m'} \tag{6}$$

for the kinematic scattering matrix

$$G_{mm'}=H_{mm'}-A_m\delta_{mm'}=K(\gamma_m\gamma_{m'})^{-1/2}\times(g_{mm'}-\alpha_m\delta_{mm'}), \quad (7)$$

where $\delta_{mm'}$ is a Kronecker delta, which is equal to zero when $m\neq m'$ and to unity when $m=m'$.

The matrix $G_{mm'}$ has a rank of $2n$. Accordingly, there are $2n$ different characteristic solutions of the problem (6), which are distinguished by the index j. Unlike other dynamical systems, the matrix $G_{mm'}$ is non-Hermitian, since the matrix $g_{mm'}$ is non-Hermitian in the general case with consideration of the absorption of x rays. However, the parts of $g_{mm'}$ which describe scattering and absorption separately are Hermitian. Nevertheless, the matrix $G_{mm'}$ is still non-Hermitian even for a nonabsorbing crystal, if among the parameters γ_m there are some which have negative values. This always occurs in the case of back diffraction. Therefore, the eigenvalues ε of the problem, i.e., the dispersion corrections to the wave vectors, are complex even for a nonabsorbing crystal. In addition, some of them have a negative imaginary part, which corresponds to growth of the Bloch waves as they move into the crystal. This, in turn, causes some difficulty in solving the boundary-value problem by numerical methods on a computer.

The general solution of the boundary-value problem in a form which is stable toward increasing Bloch waves was given in Refs. 14 and 15. Below we shall briefly formulate the solution method used. For this purpose, we order the elements in the matrix of eigenvectors B_{mj} so that the index m corresponds to decreasing values of the parameter γ_m and the index j corresponds to decreasing values of the imaginary part of the eigenvalue ε_j''. If the number of Laue beams corresponding to the passage of radiation through the crystal plate for which $\gamma_m>0$ is equal to n_L, then the number of values of m corresponding to these waves and the number of solutions with a positive imaginary part of the eigenvalue ($\varepsilon_j''>0$) are equal to the same number $2n_L$. We denote the set of such values of the indices m and j by the single index L, and we denote the set of remaining values by the single index B. This allows us to divide the complete matrix of eigenvectors B_{mj} obtained as a result of the numerical solution of (6) into the four submatrices B_{LL}, B_{LB}, B_{BL}, and B_{BB}, of which the diagonal matrices B_{LL} and B_{BB} are strictly square, and the off-diagonal matrices are rectangular in the general case. The set of amplitudes for the reflection of Laue-type plane waves ($\gamma_m>0$) into Bragg-type plane waves ($\gamma_m<0$) is described by the block M_{BL} of the complete dynamical scattering matrix.

In this paper we analyze the back reflectivity in the approximation of a thick absorbing crystal, in which the increasing Bloch waves can be completely neglected. In this case the block of the dynamical scattering matrix of interest to us is described by the simple expression

$$M_{BL}=B_{BL}(B_{LL})^{-1}. \quad (8)$$

If the incident plane wave has the index 0 and is polarized in the s state (these conditions correspond to synchrotron radiation) and if the back-reflected wave has the index $k=n-1$ and its polarization state is not analyzed, then the experimentally measured reflectivity is described by the expression

$$R_{k0}^{(s)}=\sum_{s'}|M_{ks',0s}|^2. \quad (9)$$

The parameters of the problem are the components of the vector $\mathbf{q}=\mathbf{K}_0+\mathbf{h}_k/2$, which describe small deviations of the wave vector of the incident wave from the direction corresponding to strict backscattering $\mathbf{K}_0^{(0)}=-\mathbf{h}_k/2$. It is convenient to represent the vector $\mathbf{q}$ in the form

$$\mathbf{q}=K(\theta_1\mathbf{e}_{0\pi}+\theta_2\mathbf{e}_{0\sigma}+\theta_\omega\mathbf{s}_0), \quad (10)$$

where the parameters θ_1 and θ_2 describe the angular deviations of the incident beam and $\theta_\omega=(\omega-\omega_c)/\omega_c$ describes the spectral back reflection line sought. The critical frequency is $\omega_c=c|\mathbf{h}_k|/2$, the critical wavelength is $\lambda_c=2d_k$, where d_k is the interplanar distance for the back-reflecting atomic planes, and the crystal photon energy (in keV) is $E_c=12.4/\lambda_c$, where λ_c is measured in angstroms. With consideration of (10), the parameters of the deviation from the Bragg condition in the linear approximation with respect to $\mathbf{q}$ equal

$$\alpha_m=2(\mathbf{h}_m\cdot\mathbf{q})/K^2=2K^{-1}[(\mathbf{h}_m\cdot\mathbf{e}_{0\pi})\theta_1+(\mathbf{h}_m\cdot\mathbf{e}_{0\sigma})\theta_2+(\mathbf{h}_m\cdot\mathbf{s}_0)\theta_\omega]. \quad (11)$$

In experiments the incident beam always has a finite angular divergence, and the monochromator has a finite width. For simplicity, we assume that the shape of the angular and frequency spectra of the monochromator is rectangular. Thus, the spectral reflection line interesting us can be calculated from the formula

$$\overline{R_{k0}^{(s)}}(\theta_\omega)=\frac{1}{T_\theta^2T_\omega}\int d\theta_{\omega'}\int d\theta_1 d\theta_2 R_{k0}^{(s)}(\theta_1,\theta_2,\theta_{\omega'}-\theta_\omega), \quad (12)$$

where T_θ and T_ω specify the angular and frequency widths of the monochromator, respectively, and the integration is performed in these limits.

3. (1,9,9) BACK REFLECTION IN SILICON UNDER THE CONDITIONS OF SIX-WAVE DIFFRACTION. GEOMETRY AND CALCULATION RESULTS

In crystals of cubic symmetry strict backscattering on a reciprocal-lattice vector with fairly large Miller indices is always accompanied by reflection into other reciprocal-lattice vectors, which satisfy the Bragg conditions as a consequence of the symmetry of the crystal lattice. For example, in a silicon crystal the (1,9,9) reflection occurs simultaneously with the (−4,0,4), (−4,4,0), (5,9,5), and (5,5,9) reflections, so that six-wave diffraction is realized when the Bragg conditions are strictly satisfied. In this case the truncated Bloch waves are sums of plane waves with the wave vectors $\mathbf{k}_m=\mathbf{K}_0+\mathbf{h}_m$, which have the following values in units of π/a, where a is the lattice constant, in the coordinate system of the reciprocal lattice of the crystal:

(−0.5,−4.5,−4.5); (−4.5,−4.5,−0.5);

(−4.5,−0.5,−4.5);

(4.5,4.5,0.5); (4.5,0.5,4.5); (0.5,4.5,4.5).

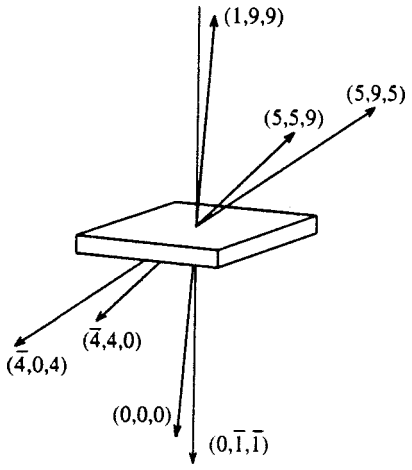

FIG. 1. Directions of the diffracted beams relative to the crystal plane. The plane of the plate is perpendicular to the $(0,-1,-1)$ direction.

Let the crystal plate be cut so that the normal to the surface is parallel to the $(0,-1,-1)$ direction. In this case the parameters γ_m are equal to 0.997, 0.554, 0.554, -0.554, -0.554, and -0.997, respectively, i.e., we have three Laue-type waves and three Bragg-type waves. Figure 1 shows the directions of the diffracted beams relative to the crystal plate. The polarization vectors in each wave can be chosen arbitrarily. Taking into account the scattering symmetry, we choose the polarization vectors in the following manner. The vector $\mathbf{e}_{0\pi}$ is parallel to the $(0,1,-1)$ direction, and the remaining vectors are defined according to the formulas

$$\mathbf{e}_{m\sigma}=\mathbf{s}_m\times\mathbf{e}_{0\pi},\quad \mathbf{e}_{m\pi}=\mathbf{e}_{m\sigma}\times\mathbf{s}_m. \tag{13}$$

At the same time, the set of three vectors $\mathbf{e}_{0\pi}$, $\mathbf{e}_{0\sigma}$, and $\mathbf{s}_0$ is used to resolve the vector $\mathbf{q}$, as follows from formula (10).

The numerical values for the Fourier components χ_m of the polarizability of the crystal were obtained using Stepanov's X0H program. When this work was carried out, this program was freely available on the Internet.[16]

As we know, a symmetric 2×2 matrix with diagonal elements equal to one another has eigenvectors with components that are equal in absolute value and thus corresponds exactly to dynamical diffraction. Therefore, the centers of the two-wave reflection maxima are determined with consideration of the dynamical displacement of the parameters by the relations $A_m=H_{mm}-H_{00}$, which depend, among other things, on the parameters γ_m, rather than by the conditions $\alpha_m=0$. In the case under consideration these conditions have the following form in microradians (μrad):

$$\begin{aligned}
&(-4,0,4):\ -0.118\theta_1-0.176\theta_2-0.105\,\theta_\omega=-0.271\\
&(-4,4,0):\ +0.118\theta_1-0.176\theta_2-0.105\,\theta_\omega=-0.271\\
&(5,9,5):\ -0.118\theta_1-0.176\theta_2+0.429\,\theta_\omega=+0.949\\
&(5,5,9):\ +0.118\theta_1-0.176\theta_2+0.429\,\theta_\omega=+0.949\\
&(1,9,9):\ +0.296\,\theta_\omega=+0.678.
\end{aligned} \tag{14}$$

As follows from these conditions, two-wave back reflection does not depend on the angular variables, but the reflection maximum is shifted with respect to the photon energy by $\Delta E=\Delta_0=E_c\theta_\omega^{(0)}=33.4$ meV.

It is difficult to graphically represent the three-dimensional dependence of the reflectivity $R_{k0}^{(s)}(\Delta E,\theta_1,\theta_2)$, where $\Delta E=E_c\theta_\omega$. Therefore, we shall present and discuss only fragments of the general dependence. Figure 2 shows the dependence of the back reflectivity for the (1,9,9) reflection in the $(\Delta E,\theta_2)$ plane of arguments at $\theta_1=0$, and Fig. 3 presents the dependence in the $(\Delta E,\theta_1)$ plane at $\theta_2=0$ for both polarization states in the incident wave. As follows from the calculations represented in these figures, the two-wave band of the back reflection maximum due to (1,9,9) diffraction vanishes as the multiwave region of parameters is approached. In addition, it is easy to discern the presence of additional reflection bands in regions where the Bragg condition for (1,9,9) diffraction is not satisfied but the Bragg condition for other reflections is satisfied.

While the value of the reflectivity increases as we move along the two-wave band of the (1,9,9) reflection from the center to the edges, it decreases as we move along the additional bands. This is because the additional reflection bands have an essentially multiwave character. The slope of these bands relative to the energy axis in the $(\Delta E,\theta_2)$ plane at $\theta_1=0$ is determined from the conditions (14). For example, the conditions for two-wave diffraction in the (5,9,5) and (5,5,9) reflections are satisfied in the line at $\theta_2=0.167\Delta E-5.932$. Here and below, the shift of the photon energy ΔE is measured in millielectron volts (meV). This means that three-wave (0,0,0; 5,9,5; 5,5,9) diffraction occurs in this line. The (1,9,9) back reflection is weak (kinematic), but it is enhanced because of the simultaneous presence of several strong waves. A more detailed analysis is given in the following section.

As follows for Fig. 2, there is a second line of additional reflection. It corresponds to three-wave $(-4,0,4;-4,4,0;1,9,9)$ diffraction. The Bragg condition for this case is obtained by subtracting the condition for the $(-4,0,4)$ reflection from the condition for the (1,9,9) reflection in formulas (14). At $\theta_1=0$ a simple calculation then permits determination of the equation of the second line at $\theta_2=-0.156\Delta E+5.932$. The two lines cross at the point $\Delta E=36.73$ meV, $\theta_2=0.742$ μrad. The lines split at the crossing point, and there is symmetry of the (1,9,9) back reflectivity in the split lines relative to the change in the sign of the quantity $\theta_2-0.742$, although the physical conditions for reflection on both sides of the symmetric pattern are different. In one case the (1,9,9) reflection is a disturbance in the Bloch wave, where the strong waves are the (0,0,0), (5,9,5), and (5,5,9) waves. In the other case the (1,9,9) wave is a strong wave together with the $(-4,0,4)$ and $(-4,4,0)$ waves, but the perturbation is a component in the incident (0,0,0) band; therefore this Bloch wave is weakly excited in the crystal. The presence of polarization in the incident wave weakly influences the two-wave band of the (1,9,9) reflection, but has a very significant effect on the additional reflection bands.

The dependence shown in Fig. 3 is even more compli-

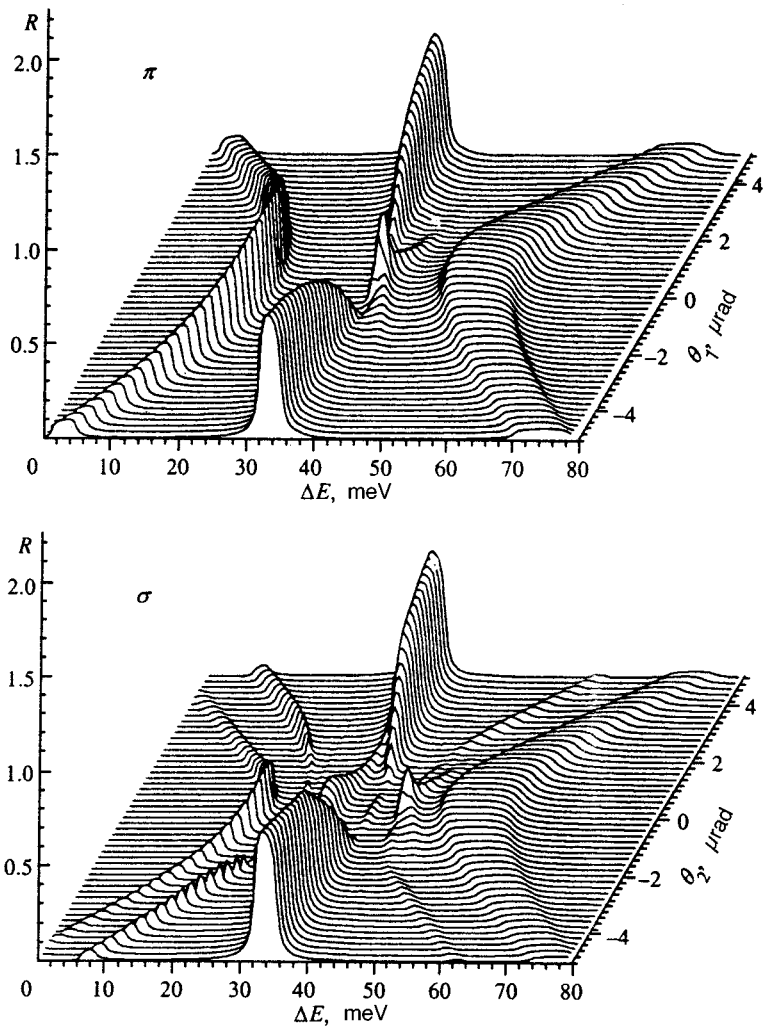

FIG. 2. Dependence of the reflectivity for (1,9,9) back reflection at $\theta_1=0$ for two polarization states of the incident wave (π and σ).

cated. The two-wave diffraction in the (0,0,0; 5,9,5) and (0,0,0; 5,5,9) reflections now takes place in the lines at $\theta_2=\pm 0.249$ ($\Delta E-32.2$). The additional reflection bands are strongly split for both polarization states. In addition, two-wave ($-4,4,0;1,9,9$) and ($-4,0,4;1,9,9$) bands are displayed in the lines at $\theta_2=\pm 0.239$ ($\Delta E-33.7$). The bands are closely spaced, although they do not coincide with one another. Therefore, the two-wave case with strong renormalization of the scattering parameters is partially realized here.

The experimental observation of the dependences of the reflectivity presented in this paper requires a strongly collimated (less than 1 μrad^2) and monochromatized (of the order of 1 meV) beam. If the beam has finite collimation and is not scanned over the angle, the dependence of the back reflectivity on the photon energy shift ΔE can be obtained by integrating over the angular variables θ_1 and θ_2 in assigned limits [see formula (12)]. Figure 4 shows the back reflection energy spectra for $T_\omega=0$ and $T_\theta=0$, 4, 8, 12, 16, 20 μrad, and ∞. The integration was carried out by simple summation on a square grid with a spacing of 0.2 μrad along both axes. For better visibility, the curves for different values of T_θ have been shifted to achieve 0.2 spacing along the vertical axis. The lower curve corresponds to $T_\theta=0$ and the upper curve (for $T_\theta=\infty$ within the approximation considered) corresponds to pure two-wave diffraction.

As follows from the calculations, multiwave effects are displayed even with collimation of the beam to 20×20 μrad^2 in the form of a lower maximum of the principal reflection and additional regions of weak reflection. However, already with angular misorientation of the beam exceeding 10×10 μrad^2, the principal maximum is fully distinguishable and has a position on the energy scale corresponding to the two-wave case. This result can be utilized in calibrating monochromators with an energy resolution of the order of 1 meV.

4. TWO-WAVE DYNAMICAL DIFFRACTION, KINEMATIC DIFFRACTION, AND THE INFLUENCE OF OTHER REFLECTIONS ON THEIR PROPERTIES

Multiwave dynamical diffraction is described by the system of equations (6), which does not have an analytic solution in the general case. Moreover, the results of detailed studies only of cases of systematic diffraction, in which all the vectors of the reciprocal lattice lie in a single plane, have been published hitherto. In such cases the variation of the energy of the incident photons leads only to variation of the reference point on the plane of angular parameters without alteration of the angular dependence of the reflectivities. The case which we considered refers to nonsystematic (random)

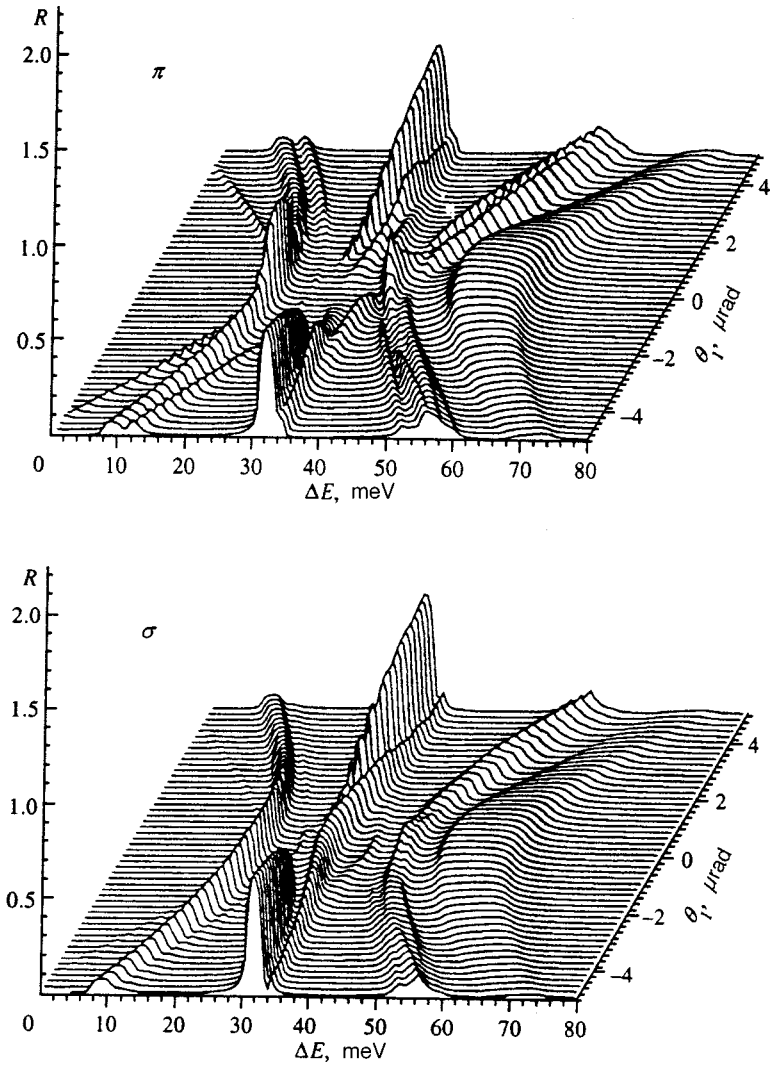

FIG. 3. Dependence of the reflectivity for (1,9,9) back reflection at $\theta_2=0$ for two polarization states of the incident wave.

diffraction in Chang's terminology.[17] Nevertheless, we can use the approximate approach previously developed for a qualitative analysis of the calculation results.

Let us consider the important special case where two of the set of parameters A_m characterizing the deviation from the Bragg conditions are close to one another, for example, the parameters with the indices i and j, while the remaining parameters have values differing strongly from these two. In this case it is natural to presume that only the components B_i and B_j of the eigenvector will have large and comparable values, while the remaining components will be small. We first consider the situation in which the small components can be completely neglected and the polarization can be separated. This corresponds to two-wave diffraction, for which in the system of equations (6) it is sufficient to retain only two equations:

$$(\varepsilon+A_i-H_{ii})B_i-H_{ij}B_j=0,$$

$$-H_{ji}B_i+(\varepsilon+A_j-H_{jj})B_j=0. \tag{15}$$

This system has two solutions, in which

$$\varepsilon_{1,2}=H_{ii}-A_i+0.5[-a\pm(a^2+4H_{ij}H_{ji})^{1/2}],$$

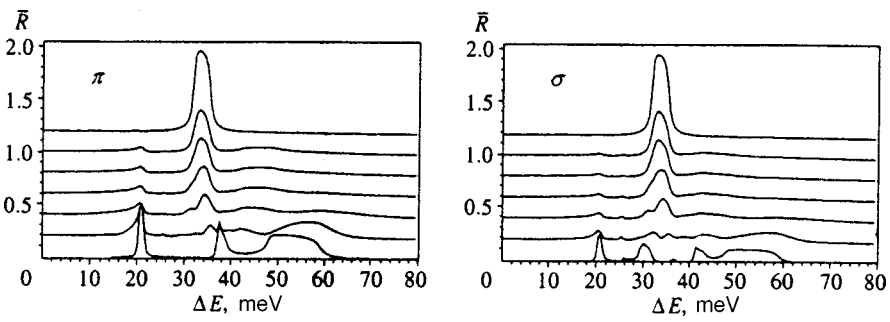

FIG. 4. Energy spectra of the angle-integrated back reflectivity for various values of the collimation of the incident beam: 0×0 (lower curve), 4×4, 8×8, 12×12, 16×16, $20\times20\,\mu\text{rad}^2$, and the two-wave case (upper curve). For better visibility the curves have been shifted along the vertical axis with 0.2 spacing.

$$a=(A_j-A_i)-(H_{jj}-H_{ii}),\ \ B_j/B_i=(\varepsilon+A_i-H_{ii})/H_{ij}. \tag{16}$$

Here the branch with a positive imaginary part is chosen for the square root.

If the index $i=0$ corresponds to the incident beam, and the index $j=h$ corresponds to the (1,9,9) back reflection, then in the approximation of a thick absorbing crystal, the reflection amplitude is exactly equal to the ratio between the components of the Bloch wave and can be written in the standard notation[1] as follows:

$$\frac{B_h}{B_0}=i\frac{p+\sqrt{p^2-4\beta\chi_h\chi_{\bar{h}}C}}{2\chi_{\bar{h}}}, \tag{17}$$

where

$$p=\alpha\beta-\chi_0(1+\beta),\quad \beta=\gamma_0/|\gamma_n|,\quad C=(\mathbf{e}_{0s}\mathbf{e}_{hs}). \tag{18}$$

Here it has been taken into account explicitly that $\gamma_h<0$. In the case of back reflection $\beta=1$. Equations (17) and (18) correspond exactly to the upper curve in Fig. 4 with consideration of the relation $\alpha=-4\theta_\omega=-4\Delta E/E_c$.

In the kinematic approximation, in which the rescattering between the weak components of the mixed Bloch wave for $m\neq i,j$ can be neglected and only the single scattering from strong waves into weak waves need be taken into account, the weak components are given by the following expression:

$$B_m=\frac{H_{mi}B_i+H_{mj}B_j}{\varepsilon+A_m-H_{mm}}. \tag{19}$$

As follows from this formula, ordinary single-wave single scattering from the incident wave with $i=(000)$ into the back reflection wave with $m=(1,9,9)$ can be enhanced in the presence of several strong waves, and the appearance of additional reflection maxima can be expected in the case where the Bragg condition is satisfied for some wave with $j\neq(1,9,9)$ and this wave is scattered in phase with the incident wave.

In addition, we can write the system of equations for strong waves in the more exact form

$$\begin{aligned}(\varepsilon+A_i-H_{ii})B_i-H_{ij}B_j&=\sum_m H_{im}B_m,\\ -H_{ji}B_i+(\varepsilon+A_j-H_{jj})B_j&=\sum_m H_{jm}B_m,\end{aligned} \tag{20}$$

where $m\neq i,j$ in the sum. Now, using formula (19 for weak waves and substituting it into (20), we obtain a system of the type (15), but with renormalized coefficients:

$$\begin{aligned}(\varepsilon+A_i-F_{ii})B_i-F_{ij}B_j&=0,\\ -F_{ji}B_i+(\varepsilon+A_j-F_{jj})B_j&=0,\end{aligned} \tag{21}$$

where

$$F_{kl}=H_{kl}+\sum_{m\neq i,j}\frac{H_{km}H_{ml}}{\varepsilon+A_m-H_{mm}},\quad k=i,j,\quad l=i,j. \tag{22}$$

Formulas like (19), (21), and (22) were obtained for a more general case with consideration of the polarization multipliers in Ref. 18 as a method for approximate solution of the problem. In Refs. 19–21 the mechanism for renormalization of the parameters was called virtual Bragg scattering. The same approach was used in Refs. 22 and 23 to investigate standing x-ray waves and total reflection in a forbidden reflection.

Under ordinary conditions for two-wave diffraction the parameter describing the deviation from the Bragg condition for a weak wave $A_m\gg\varepsilon$, H_{mm}, and the renormalization of the coefficient is very small. However, in a situation which is close to the pure multiwave situation, this renormalization is significant and can significantly distort the character of two-wave diffraction, i.e., can significantly shift the position of the maximum and alter its width and height. In this case even the magnitude of the dispersion correction should be calculated self-consistently.

Thus, an analysis of the multiwave corrections to the two-wave (1,9,9) back reflection reveals effects of two types. First, the two-wave back reflectivity peak is distorted as a result of the renormalization of the parameters of the scattering associated with rescattering on other reflections. Second, renormalization of the kinematic diffraction appears when the Bragg conditions for any of the other reflections are satisfied with consideration of the renormalization of its parameters. Significant interference of the two scattering channels then occurs, as a result of which, as the analysis showed, the single-band approximation does not provide the required accuracy in comparison to an exact multiwave calculation. Therefore, the formulas presented in this section are suitable only for a qualitative understanding of the results of the exact calculation presented in Figs. 2 and 3, but cannot be used directly for calculations.

The mechanism discussed here can also be considered in the case where the conditions for three-wave diffraction are satisfied simultaneously, as occurs at $\theta_1=0$. The formulas presented above can easily be generalized to this case. The situation is far more complicated when the regions of two-wave diffraction for different reflections are fairly close, but do not coincide exactly. In this case, the interference of different scattering channels leads to a complicated structure of peaks of kinematic scattering, as is clearly seen in Fig. 3.

We express our thanks to Yu. Shvyd'ko for formulating the problem and taking an interest in this work.

*)E-mail: kohn@kurm.polyn.kiae.su

[1] Z. G. Pinsker, *X-Ray Crystal Optics* [in Russian], Nauka, Moscow (1982).
[2] K. Kohra and T. Matsushita, Z. Naturforsch. A **27**, 484 (1972).
[3] O. Brümmer, H. R. Höche, and J. Nieber, Phys. Status Solidi A **53**, 565 (1979).
[4] A. Caticha and S. Caticha-Ellis, Phys. Rev. B **25**, 971 (1982).
[5] W. Graeff and G. Materlik, Nucl. Instrum. Methods Phys. Res. **195**, 97 (1982).
[6] V. I. Kushnir and É. V. Suvorov, JETP Lett. **44**, 262 (1986).
[7] R. Verbeni, F. Sette, M. H. Krisch, U. Bergmann, B. Gorges, C. Halcoussis, K. Martel, C. Masciovecchio, J. F. Ribois, G. Ruocco, and H. Sinn, J. Synchrotron Radiat. **3**, 62 (1996).
[8] A. Caticha, K. Aliberti, and S. Caticha-Ellis, Rev. Sci. Instrum. **67**, 1 (1996).

[9] Yu. V. Shvyd'ko, E. Gerdau, J. Jäschke, O. Leupold, M. Lucht, and H. Ruter, Phys. Rev. B **57**, 4968 (1998).
[10] Yu. V. Shvyd'ko *et al.*, submitted to Hyperfine Interactions (1999).
[11] A. I. Chumakov, R. Rüffer, A. O. R. Baron, H. Grünsteudel, H. F. Grünsteudel, and V. G. Kohn, Phys. Rev. B **56**, 10 758 (1997).
[12] V. G. Kohn, A. I. Chumakov, and R. Rüffer, Phys. Rev. B **58**, 8437 (1998).
[13] V. G. Kohn, Phys. Status Solidi A **54**, 375 (1979).
[14] V. G. Kohn, J. Mosc. Phys. Soc. **1**, 425 (1991).
[15] V. G. Kohn, Zh. Éksp. Teor. Fiz. **105**, 665 (1994) [JETP **78**, 357 (1994)].
[16] S. A. Stepanov, http://sergey.bio.aps.anl.gov.
[17] S. L. Chang, *Multiple Diffraction of X Rays in Crystals*, Springer-Verlag, Berlin (1984) [Russ. transl., Mir, Moscow (1987)].
[18] R. Hoier and K. Martinsen, Acta Crystallogr. A **25**, 854 (1983).
[19] L. D. Chapmann, D. R. Yoder, and R. Colella, Phys. Rev. Lett. **46**, 1578 (1981).
[20] R. Colella, Z. Naturforsch. A **37**, 437 (1982).
[21] M. C. Schmidt and R. Colella, Phys. Rev. Lett. **55**, 715 (1985).
[22] V. G. Kon, Kristallografiya **33**, 567 (1988) [Sov. Phys. Crystallogr. **33**, 333 (1988)].
[23] V. G. Kohn, Phys. Status Solidi A **106**, 31 (1988).

Translated by P. Shelnitz

Phase transformations in the disordered antiferromagnetic Ising model

V. V. Prudnikov,*) O. N. Markov, and E. V. Osintsev

Omsk State University, 644077 Omsk, Russia
(Submitted 23 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 953–961 (September 1999)

We employ the Monte Carlo method to provide a computer description of the thermodynamic properties of the disordered antiferromagnetic Ising model for a simple cubic lattice with allowance for the nearest neighbors interaction and the next-nearest neighbors interaction. We construct the phase diagram for the model in an external magnetic field for systems with spin concentrations $p=1.0$, 0.95, and 0.8. We also establish the tricritical behavior of the model. Finally, we study the critical behavior of the model in weak fields. © *1999 American Institute of Physics.* [S1063-7761(99)01409-2]

In the last two decades the efforts of many researches have been focused on understanding of how impurities and other defects affect the behavior of various systems in phase transitions. Of special interest here is the effect of frozen impurities, whose presence manifests itself in the form of random local perturbations of temperature in ferromagnets and antiferromagnets in the absence of an external field or in the form of random magnetic fields for antiferromagnets in a uniform magnetic field. In view of the fact that a magnetic field breaks the symmetry of the system with respect to a change in the signs of spins, the statistical properties of such disordered systems differ substantially.

Studies have shown[1] that in phase transitions without an external magnetic field the presence of frozen impurities alters the properties of only those magnetic substances whose specific heat in the homogeneous state diverges at the critical point. In all other cases the impurities have no effect on the behavior of the magnetic substances at the critical point. The given criterion is met only by systems whose effective Hamiltonian near the point is isomorphic to the Ising model. The literature devoted to study of the critical properties of the Ising model is vast (see, e.g., the review articles in Refs. 2 and 3). For dilute systems, good agreement has been achieved between the theoretical results and the results of experiments and Monte Carlo simulations.

As for magnetic substances with random fields, notwithstanding the intensive theoretical and experimental research done in the last two decades,[3] there are still very few reliable facts characterizing the behavior of such systems. In particular, the nature of the phase transition in the random-field Ising model is yet to be determined, while the results obtained by computer simulation of such systems are contradictory. The almost single reliable fact established so far is that the upper critical dimensionality for this phase transition (the dimensionality of the system above which critical phenomena are described by the mean-field theory) is six,[3] in contrast to the value of four for homogeneous systems. Lately, in the problem of the lower critical dimensionality of the transition in the random-field Ising model (the dimensionality d_l of the system above which long-range ordering sets in at finite temperatures), where Imry and Ma[4] argued that $d_l=2$ and Parisi and Sourlas[5] that $d_l=3$, Imbrie[6] and Bricmont and Kupiainen[7] came to the conclusion that $d_l=2$. However, the nature of the phase transition in the three-dimensional Ising model is still unclear. According to Young and Nauenberg[8] and Rieger and Young[9], this is a first-order phase transition down to very random-field values, while according to Ogielski and Huse,[10,11] it is a second-order phase transition.

Two (qualitatively) equivalent models have been used to describe the effect of random fields on the behavior of magnetic systems: the model of random-field Ising ferromagnets (RFIM)[12] and the model of Ising dilute antiferromagnets in a field (DAFF).[13] Real magnetic systems with random-field effects are antiferromagnets with frozen impurities of nonmagnetic atoms. Such systems exhibit not only effects associated with the antiferromagnetic interaction of nearest neighbors but also effects associated with the ferromagnetic interaction of next-nearest neighbors. The DAFF model does not account for the competition of the ferromagnetic interaction, with the result that its real use (like that of the RFIM model) is fairly limited.

To establish the features of the phase transitions in random-field magnetic substances that set these substances apart from systems with a random local temperature (random spin interactions), we did a Monte Carlo simulation of the critical behavior of an Ising dilute antiferromagnet using a simple cubic lattice with allowance for the nearest-neighbor interaction and the next-nearest-neighbor interaction. The model Hamiltonian is

$$\mathscr{H}=J_1\sum_{i,j}p_ip_j\sigma_i\sigma_j+J_2\sum_{i,k}p_ip_k\sigma_i\sigma_k+\mu H\sum_i\sigma_i, \tag{1}$$

where $\sigma_i=\pm 1$, $J_1=1$ characterizes the exchange interaction between the nearest-neighbor spins (which is of an antiferromagnetic nature), $J_2=-1/2$ characterizes the ferromagnetic interaction of the next-nearest-neighbor spins, H is the strength of the uniform external magnetic field, and p_i and p_j are the random variables described by the distribution function

$$P(p_i)=p\,\delta(p_i-1)+(1-p)\,\delta(p_i) \tag{2}$$

and characterizing the frozen nonmagnetic impurity atoms (vacant sites) distributed over the lattice sites with a concentration $c_{\rm imp}=1-p$. The given model with $p=1.0$ and competing interactions has been studied for more than two decades by Monte Carlo methods.[14,15] However, we are the first to use the model to describe the influence of disorder effects on the critical behavior of systems. From the viewpoint of physics, this model is the most realistic one. The size of the random-field effects in this model and in real magnetic systems is determined by the impurity concentration and the magnitude of the external field. For this reason the parameters of the model uniquely correspond to the parameters of a real physical experiment. The situation is different in the RFIM model: the magnitude of the random field is fixed and cannot be unambiguously related to the parameters of a real physical experiment, i.e., the impurity concentration in the sample and the magnitude of the external field.

What makes the present disordered model so remarkable is that at $H=0$ it is able to describe the critical behavior of a system with random spin interactions, while for $H<H_t$, according to the results discussed below, it demonstrates the critical behavior of a system with random fields. For $H>H_t$, magnetization fluctuations violate the stability of the second-order phase transition, so that phase transformations in the system acquire the features of a first-order phase transition. At $H=H_t$ and $T=T_t$ the system exhibits tricritical behavior. Thus, the given model describes a broad class of phase transitions and makes it possible to study the effect of the disorder introduced by impurities on the thermodynamic characteristics of the system in phase transitions.

To establish the features of the thermodynamic characteristics of a disordered system that determine the system's behavior for different types of phase transitions, we must first construct the phase diagram of the system, which specifies the dependence of the phase transition temperature $T_{\rm ph}$ on the magnetic field strength H at a given spin concentration p, i.e., $T_{\rm ph}=T_{\rm ph}(H,p)$. In finding the component of the phase diagram $T_c=T_c(H,p)$ corresponding to second-order phase transitions we must bear in mind that the critical behavior of the antiferromagnetic system is determined by strong, long-lived fluctuations of the ''staggered'' susceptibility $M_{\rm stg}$, the difference of the sublattice magnetizations. The measure of magnetic fluctuations is the linear size $\xi(T)$ of a characteristic magnetic domain, a region with strongly correlated spins. As T approaches T_c, the correlation in the spin orientations grows in strength and the increase in $\xi(T)$ is described by a power law with an exponent ν: $\xi(T)\propto|T-T_c|^{-\nu}$. The ''staggered'' susceptibility and the specific heat of the system also experience an anomalous increase near T_c: $\chi_{\rm stg}\propto|T-T_c|^{-\gamma}$ and $C(T)\propto|T-T_c|^{-\alpha}$, where α and γ are the critical exponents. Due to the long-lived fluctuations of $M_{\rm stg}$, the relaxation time τ_p also increases without limit near T_c. Such behavior of the thermodynamic functions and the physical parameters is observed in the immediate vicinity of T_c, i.e., $(T-T_c)/T_c\ll 1$, for systems that are assumed to be at the thermodynamic limit (the number N of particles in the system tends to infinity and so does the system volume V, but $N/V\to$const). In a finite system there can be no real second-order phase transition. Nevertheless, it is believed that if $\xi(T)$ is smaller than the linear size of the system, a finite system correctly describes the properties of an infinite system. In other words, if T is not too close to T_c, model calculations must yield results comparable to those for an infinite system. To find T_c, we use the assumption that the critical properties for systems of different scales L are attained only as $L\to\infty$.

One method that can be used to find the critical temperature T_c in an infinite system ($L=\infty$) is to determine the ''critical temperature'' of a finite system $T_c(L)$ from the position of the peak in the temperature dependence of the ''staggered'' susceptibility $\chi_{\rm stg}(T,L)$ and use the scaling asymptotic dependence

$$T_c(L)-T_c(L=\infty)\simeq aL^{-1/\nu},\quad L\gg 1, \tag{3}$$

where a is a constant depending on the details of the model and the boundary conditions. However, in the case of computer simulation of the critical behavior of homogeneous system, the best method so far for determining $T_c(L=\infty)$ is Binder's method of cumulants,[16] which in our case involves introducing the cumulant U of the form,

$$U=\frac{1}{2}\left(3-\frac{[\langle M_{\rm stg}^4\rangle]}{[\langle M_{\rm stg}^2\rangle]^2}\right), \tag{4}$$

where the angle brackets stand for statistical averaging and the square brackets, for averaging over the different impurity configurations. The cumulant $U(L,T)$ has a scaling form

$$U(L,T)=u(L^{1/\nu}(T-T_c)) \tag{5}$$

that is important for describing the behavior of finite systems and does not exhibit a multiplicative dependence on L. The cumulant is defined in so as to obey $0\leqslant U\leqslant 1$ and so that at temperatures above T_c it tends to zero as $L\to\infty$. The scaling dependence of the cumulant makes it possible to determine the critical temperature $T_c(L=\infty)$ from the coordinate of the point of intersection of the curves specifying the temperature dependence $U(L,T)$ for different values L. More than that, it can easily be shown that in the critical region, as $T\to T_c$,

$$\frac{dU}{dT}=aL^{1/\nu}(1+bL^{-\omega}), \tag{6}$$

with the result that the exponent ν can be found from the maximum slope of the cumulants corresponding to different values of L in the limit $L\to\infty$ near their intersection point.

We examined cubic lattices with $L=12$, 18, 24, and 32 with spin concentrations $p=1.0$, 0.95, and 0.8. In growing the impurity configurations, the number of impurity atoms (vacant sites) $(1-p)L^3$ was taken to be the same for each antiferromagnetic lattice and a procedure for sampling their random distributions among the sites of the sublattices was implemented. The initial localization of the ''critical temperature'' of the finite system, $T_c(L,H)$, was determined in relation to the maximum in the temperature dependence of the ''staggered'' susceptibility $\chi_{\rm stg}(T,L)$ for the following values of field strength (in units of J_1/μ): $H=0$, 1, 2, 3, 4, 4.5, and 5.2 (Fig. 1 depicts the dependence of $\chi_{\rm stg}$ on T for

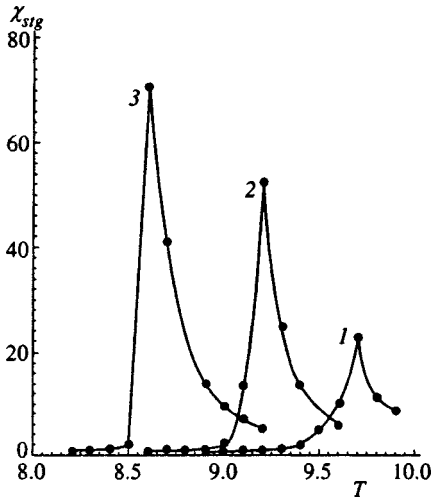

FIG. 1. Temperature dependence of the ''staggered'' susceptibility χ_{stg} along the curves of second-order phase transition in the system 18^3 with $p=0.95$ at $H=0$ (curve *1*), $H=2$ (curve *2*), and $H=3$ (curve *3*).

three values of H). Then we employed the method of cumulants to determine $T_c(L=\infty,H)$ more accurately.

To obtain reliable values off the equilibrium thermodynamic characteristics of the behavior of systems in the critical region, the procedures of statistical averaging and averaging over the different impurity configurations must be carried out after the specific system has reached the state of equilibrium. The critical behavior of various systems, especially disordered, is characterized by anomalously long relaxation times, which increase substantially as the system being modeled grows in size.

Disordered systems with random fields, such as spin glasses, belong to the class of frustrated systems, in which frustration is due to their structural disorder. In frustrated systems, the competing interactions with different spin ordering contradict each other and cause a series of events in which there is not a single one of the possible spin configurations that can minimize all the components of the system Hamiltonian simultaneously. In disordered antiferromagnets the aligning of the spins in a uniform external field is the cause for competition with antiferromagnetic ordering in the low-temperature range.

Some of the researchers[13,17–21] who studied the behavior of the RFIM and DAFF models with random-field effects found that at low temperature a set of metastable states separated by energy barriers appears. One such state is a state with long-range order (ferromagnetic for RFIM and antiferromagnetic for DAFF), while the other states have different configurations of the domain structure. It was shown that a system initially frozen in one of its domain states anomalously slowly relaxes to the state with long-range order. The effect of various experimental conditions (the freezing and heating rates for the system in the presence or absence of an external field) on the nature of the states that occur in the ordered phase and the size of the irreversible effects as functions of the random-field amplitude and the proximity to the critical temperature was studied in Refs. 13, 17, 20, and 21.

In our studies of the relaxation properties of the model with $L=24$ and a spin concentration $p=0.8$ we used a Monte Carlo simulation and found that in the low-temperature phase the system usually relaxes in the course of 2000 to 4000 steps per spin to one of its metastable states, and this process is followed by fluctuation flips of the ''staggered'' magnetization with intervals amounting, on the average, to 5000 steps. These flips are accompanied by an anomalously slow relaxation of the values of ''staggered'' magnetization through a series of metastable states to a state close to equilibrium in the course of 100 000 steps per spin.

To guarantee that the system reaches an equilibrium state at temperatures close to critical and to determine the thermodynamic characteristics of such states, we carried out a procedure of slow quasistatic freezing of the system from the disordered phase, beginning at a temperature at which no metastable states were found in all the ''runs.'' The procedure of quasistatic freezing amounted to a relaxation regime consisting of 3000 steps, repeated at each temperature, followed by an averaging regime consisting of 10 000 steps and a lowering of the temperature with a step $\Delta T=0.01$, with each initial spin configuration being the one obtained at the last step of the preceding temperature. To avoid possible irreversible effects, especially in strong fields,[13] which manifest themselves in the difference of the thermodynamic quantities in the thermal cycling from the disordered phase to the ordered phase and back, each ''run'' consisted of the procedure of quasistatic freezing described above followed by heating.

In the process of calculating the cumulants $U(L,T)$, for each lattice of size L and fixed H and p we did statistical averaging over ten ''runs'' with different initial spin configurations for each impurity configuration, which was followed by averaging over 20 to 40 different impurity configurations. Note that using cumulants provides a good test of the type of transition in the system. For instance, in the case of second-order phase transitions, the curves representing the temperature dependence of the cumulates have an intersection point (actually a triangle if a finer scale is used), as shown in Fig. 2, while in the case of first-order phase transitions the temperature curves of the cumulants have a specific shape and do not intersect. Studies have revealed that as the impurity concentration and the magnetic field strength increase, so does the error in determining the average value of the critical temperature (the area of the triangle of intersection of the temperature curves of the cumulants increases; see Fig. 2), due to the increase in strength of the configurational impurity fluctuations.

The tricritical point $[T_t(L),H_t(L)]$ was determined for each lattice with $L=12$, 18, 24, and 32 and spin concentrations $p=1.0$, 0.95, and 0.8 from the behavior of the isothermal magnetization M under variations of the field strength H. The change of the order of the phase transition from first to second at the tricritical point was found to be accompanied by the disappearance of the hysteresis loop, which characterizes the dependence of M on H with a decrease or increase of field strength H along the curve of first-order transitions (Fig. 3). The value of the tricritical temperature T_t for an infinite system was determined by extrapolating $T_t(L)$ as $L\to\infty$ according to (3), while the corresponding value of

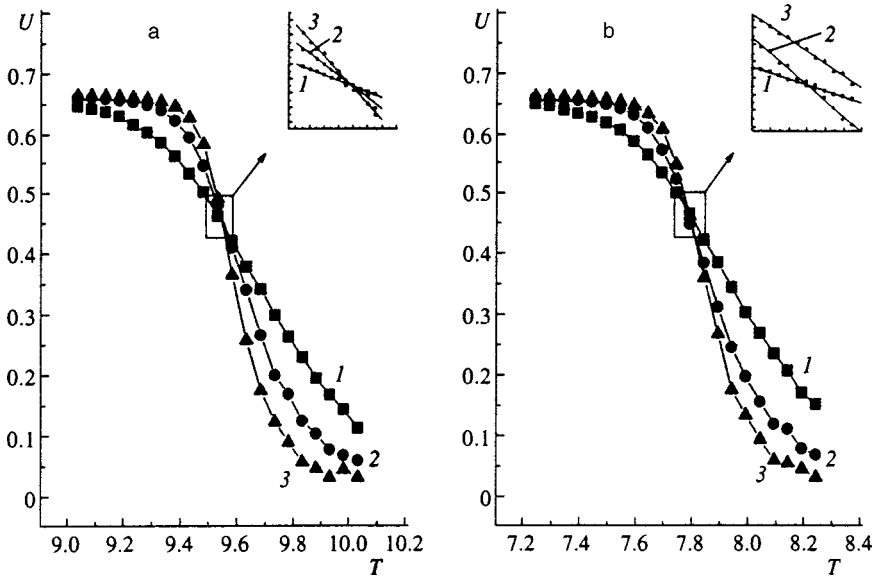

FIG. 2. Curves representing the temperature dependence of the cumulants $U(T,H)$ for lattices with $L=18$ (curves *1*), $L=24$ (curves *2*), $L=32$ (curves *3*) and spin concentrations $p=0.95$ (a) and $p=0.8$ (b) at $H=1$.

H_t was found by extrapolating $H_t(L)$ in accordance with the scaling relationship

$$H_t(L)-H_t(L=\infty)\simeq bL^{-(d+2-\eta_t)/2}, \tag{7}$$

where d is the dimensionality of the system, and $\eta_t \ll 1$ is the Fisher exponent.

As a result of our investigations we arrive at the following values of the parameters that determine the critical point: $T_t=6.14\pm0.03$ and $H_t=5.40\pm0.10$ at $p=1.0$, $T_t=5.15\pm0.10$ and $H_t=5.35\pm0.07$ at $p=0.95$, and $T_t=2.64\pm0.03$ and $H_t=4.71\pm0.05$ at $p=0.8$ (here the temperature values are given in units of J_1/k, where k is Boltzmann's constant).

Note that for a homogeneous system ($p=1.0$) the value of T_t and H_t are a refinement of the corresponding values obtained by Landau,[22] because we were dealing with larger systems (in Ref. 22, $6\leqslant L\leqslant 20$).

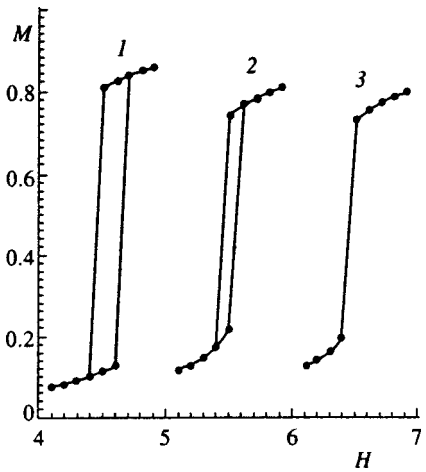

FIG. 3. Magnetization M as a function of the magnetic field strength H in the system 18^3 with $p=0.95$ at three temperature near the tricritical point $T_t=4.6$: $T=4.0$ (curve *1*), $T=4.5$ (curve *2*), and $T=4.6$ (curve *3*). The curves *1* and *3* are shifted to the left and right along the horizontal axis by -1.0 and $+1.0$, respectively.

The curves for first-order phase transitions were localized by analyzing the temperature and field dependences of the magnetization, internal energy, and specific heat.

Our investigation resulted in determining the phase diagrams (see Fig. 4) for the antiferromagnetic homogeneous Ising model ($p=1.0$), a weakly disordered model ($p=0.95$), and a highly disordered model ($p=0.8$). We see that as the impurity concentration grows, the curves representing the phase transitions shift to the region of lower temperatures and weaker magnetic fields. The localization of these curves makes it possible to carry out a detailed analysis of the special features in the critical behavior of systems with random spin interactions and random fields.

By analyzing the asymptotic scaling dependence of the cumulants for lattices with $L=12$, 18, 24, and 32 in accordance with (6) and the ''staggered'' susceptibility at the critical temperature $T_c(L=\infty,H)$ in accordance with the expression $\chi_{\text{stg}}\propto L^{\gamma/\nu}$ we found the critical exponents ν and γ for homogeneous and disordered states of the antiferromagnetic Ising model. For instance, for the homogeneous system with $p=1.0$, $H=0$, and $T_c=10.15$ we found that $\nu=0.63\pm0.01$ and $\gamma=1.25\pm0.02$, which are in good agreement with the

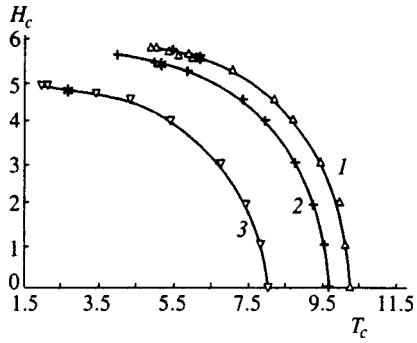

FIG. 4. Phase diagrams for the antiferromagnetic Ising model with $p=1$ (curve *1*), $p=0.95$ (curve *2*), $p=0.8$ (curve *3*); the $*$ indicate the tricritical points.

results of theoretical calculations, the high-temperature expansion, and the experiment conducted by Le Guillou and Zinn-Justin.[23] For disordered systems with random spin interactions at $p=0.95$, $H=0$, and $T_c=9.62$ and at $p=0.8$, $H=0$, and $T_c=7.97$, the resulting values of the critical exponents, $\nu=0.65\pm0.02$ and $\gamma=1.27\pm0.03$, and $\nu=0.68\pm0.02$ and $\gamma=1.31\pm0.03$ agree with Heuer's results[24] of a computer simulation of the disordered ferromagnetic Ising model and with the results of Mayer's theoretical calculation[25] for the weakly disordered Ising model. For disordered systems with random-field effects at $p=0.95$, $H=1$, and $T_c=9.53$ and at $p=0.8$, $H=1$, and $T_c=7.84$, the resulting values of the critical exponents, $\nu=0.68\pm0.02$ and $\gamma=1.35\pm0.03$, and $\nu=0.79\pm0.03$ and $\gamma=1.45\pm0.04$ demonstrate a sharp increase with the degree of disorder in the system. In contrast to a computer simulation of the RFIM model done by Young and Nauenberg,[8] in our model the values of critical exponents do not exhibit a quasi-two-dimensional nature. Such features could appear at higher values of H, where the random-field effects play an important role, but this is a problem for future investigations of our model.

To summarize, we note that our results of computer simulation of a three-dimensional disordered Ising model with antiferromagnetic interaction between the nearest neighbors and ferromagnetic interaction between the next-nearest neighbors in an external field are of primary importance. We have been the first to construct the phase diagrams for the given model at spin concentrations $p=0.95$ and 0.8. We also have convincingly shown that in weak external fields the random-field effects do not destroy the second-order phase transition. In order to localize the critical temperatures along the phase-transition curve we have employed the method of cumulants, which has made it possible to use scaling analysis not only to determine, with a high accuracy, the temperature of the second-order transition but also to test the nature of the phase transition. We have been the first to localize the tricritical points in the phase diagrams of the disordered systems under investigation by analyzing the hysteresis effects in the behavior of the magnetization. Finally, the high accuracy of localization of the critical temperatures has made it possible to find the critical exponents for the Ising model with random-temperature effects (in the absence of an external field) and random-field effects (in a finite magnetic field).

This work was sponsored by the Russian Fund for Fundamental Research (Grant No. 97-02-16124).

*)E-mail: prudnikv@univer.omsk.ru

[1] A. B. Harris, J. Phys. C **7**, 1671 (1974).
[2] R. B. Stinchcombe, in *Phase Transitions and Critical Phenomena*, Vol. 6, C. Domb and M. S. Green (Eds.), Academic Press, New York (1976), p. 151.
[3] V. S. Dotsenko, Usp. Fiz. Nauk **165**, 481 (1995) [Phys. Usp. **38**, 457 (1997)].
[4] Y. Imry and S-k. Ma, Phys. Rev. Lett. **35**, 1399 (1975).
[5] G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979).
[6] J. Z. Imbrie, Phys. Rev. Lett. **53**, 1747 (1984).
[7] J. Bricmont and A. Kupiainen, Phys. Rev. Lett. **59**, 1829 (1987).
[8] A. P. Young and M. Nauenberg, Phys. Rev. Lett. **54**, 2429 (1985).
[9] H. Rieger and A. P. Young, J. Phys. A **26**, 5279 (1993).
[10] A. T. Ogielski and D. A. Huse, Phys. Rev. Lett. **56**, 1298 (1986).
[11] A. T. Ogielski, Phys. Rev. Lett. **57**, 1251 (1986).
[12] D. P. Belanger and A. P. Young, J. Magn. Magn. Mater. **100**, 272 (1991).
[13] G. S. Grest, C. M. Soukoulis, and K. Levin, Phys. Rev. B **33**, 7659 (1986).
[14] D. P. Landau, Phys. Rev. Lett. **28**, 449 (1972).
[15] H. Müller-Krumbhaar and D. P. Landau, Phys. Rev. B **14**, 2014 (1976).
[16] K. Binder, Z. Phys. B: Condens. Matter **43**, 119 (1981).
[17] D. Andelman, H. Orland, and L. Wijewardhana, Phys. Rev. Lett. **52**, 145 (1984).
[18] G. Grinstein, and J. F. Fernandez, Phys. Rev. B **29**, 6389 (1984).
[19] H. Yoshizawa and D. P. Belanger, Phys. Rev. B **30**, 5220 (1984).
[20] D. Stauffer, C. Hartzstein, K. Binder, and A. Aharony, Z. Phys. B: Condens. Matter **55**, 352 (1984).
[21] D. Andelman and J. F. Joanny, Phys. Rev. B **32**, 4818 (1985).
[22] D. P. Landau, Phys. Rev. B **14**, 4054 (1976).
[23] J. C. Le Guillou and J. Zinn-Justin, Phys. Rev. B **21**, 3976 (1980).
[24] H.-O. Heuer, Phys. Rev. B **42**, 6476 (1990).
[25] I. O. Mayer, J. Phys. A **22**, 2815 (1989).

Translated by Eugene Yankovsky

Emergence of spectral line jumps and spectral diffusion in the two-photon correlator of a single impurity center

I. S. Osad'ko*)

Moscow State Pedagogical University, 119882 Moscow, Russia
(Submitted 26 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 962–985 (September 1999)

A dynamic theory of two-photon correlators for a single impurity center developed recently by the author [I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **113**, 1606 (1998) [JETP **86**, 875 (1998)] has been generalized to the case in which the center interacts with nonequilibrium two-level systems (TLSs) in polymers and glasses. Quantum tunneling transitions in TLS manifest themselves as random jumps of a spectral line of an impurity center. These jumps can be either spontaneous or light-induced. Interaction between the impurity center and many nonequilibrium TLSs, which exist in polymers, results in a time dependence of the optical dephasing rate $1/T_2$ of an impurity molecule, i.e., in spectral diffusion. This paper describes how the jumps of the spectral line manifest themselves in the two-photon correlator, which can be measured in experiments. © *1999 American Institute of Physics.* [S1063-7761(99)01509-7]

1. INTRODUCTION

The spectroscopy of a single molecule embedded in a polymer or glass offers unique opportunities for studying the local dynamics of these solids.[1] It is now clear that in glasses or polymers, there is not only a distribution of frequencies of electronic transitions, which determines the inhomogeneous broadening of impurity spectral lines, but also distributions of the magnitude and type of local electron-phonon coupling and of the impurity interaction with two-level systems (TLSs) in polymers and glasses.[2–4] Slow relaxation due to tunneling transitions in TLS has been extensively studied recently using the technique of single-molecule spectroscopy.

Excitations of TLSs (tunnelons) manifest themselves in electronic absorption spectra as spectral lines due to the electron-tunnelon coupling. The spectroscopy of single molecules offers unique opportunities for studying this interaction, as well as the local electron-phonon coupling.

The basic experimental technique in the spectroscopy of single molecules is excitation of fluorescence. Direct measurements of light absorption are very inefficient because it is difficult to detect absorption of a single photon from a great number of photons in the exciting light beam. In the fluorescence excitation technique, the absorption intensity is estimated on the basis of the number of emitted photons whose frequencies shifted to the red side with the respect to the excitation frequency. For this reason, the exciting radiation does not interfere with the photons emitted by a molecule.

In studies of single molecules, cw lasers with spectral widths of several megahertz are used. Under cw excitation, a molecule performs jumps between the ground and excited electronic levels at random moments of time. Therefore a light detector receives a sequence of photons emitted by the molecule and separated by random time intervals.

In single-photon detection techniques the total number of photons emitted by a molecule is counted. If the laser frequency ω_0 equals that of the spectral line maximum Ω, the light absorption intensity is maximal, hence the mean interval between emitted photons is minimal. The mean interval between emitted photons decreases with the detuning $\omega_0-\Omega$. Therefore the number of photons $N(\omega_0-\Omega)$ emitted by a molecule over the time interval t is a function of the detuning. This function describes the absorption line.

At an excitation intensity of 10^4-10^5 photons per second, which is typical of the single-molecule spectroscopy, several tens of seconds are needed to record the absorption line with good accuracy by detecting emitted photons. This shortcoming is inessential for solids which do not contain TLSs. But in polymers, where tunneling transitions in TLSs occur, one has to deal with jumps of spectral lines, after which excitation of a molecule, hence emission of photons, terminates. Such jumps on the frequency scale are shown in Fig. 1. As will be demonstrated below, these jumps of spectral lines are due to interaction between molecules and nonequilibrium TLSs, which are always present in a polymer or glass. It is clear on the intuitive level that the frequency of such spectral jumps should reflect the relaxation rate of TLSs. Unfortunately, the one-photon technique based on counting the total number of emitted photons and described above is not a convenient tool for measuring the TLS relaxation rate. The TLS relaxation time, however, can be easily assessed using the technique of two-photon correlators.[1,4,5]

In two-photon detection techniques, pairs of photons emitted by a molecule with a certain time delay t between the two photons of the pair are counted. Figure 2 shows three such pairs corresponding to one t. The count rate $p(t)$ of such pairs is called the two-photon correlator. It is also a function of the incident light frequency, i.e., $p(t)=p(\omega_0-\Omega,t)$. The task of the theoretical research is determination of a mathematical formula for $p(t)$. A quantum mechanical theory of two-photon correlators for a three-level molecule,

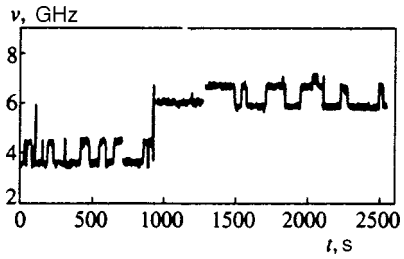

FIG. 1. Jumps of the 580.77-nm spectral line of terrylene molecule doped polyethylene.[3]

whose density operator was analyzed using optical Bloch equations or three kinetic equations, was described in detail in a previous publication.[5] It turned out that

$$p(t)=\frac{\rho_1(t)}{T_1}, \tag{1}$$

where $\rho_1(t)$ is the probability of exciting a molecule derived from the Bloch equations. The probability $\rho_1(t)$ calculated in this manner,[5] however, did not take account of the presence of nonequilibrium TLSs in a solid. A theory without this limitation will be developed in this paper, and this will be a dynamic theory of the two-photon correlator taking into account interaction with nonequilibrium TLSs whose relaxation times range from fractions of microseconds to several hours or even weeks.

The paper is organized as follows. Section 2 discusses the full Hamiltonian of the electron-photon-phonon-tunnelon system and interactions between its components. In Sec. 3, on the basis of this Hamiltonian, an infinite set of equations for the full density matrix of this system will be derived, and basic approximations will be discussed, which allow one to simplify these equations and replace them with four kinetic equations for populations of the states of a quantum system consisting of a two-level molecule coupled to phonons and one TLS. In the process of realizing these approximations, microscopic expressions for both tunneling transition rates of TLS and coefficients of light absorption and emission by a two-level molecule interacting with phonons and one nonequilibrium TLS will be derived concurrently. This transition from the infinite set of equations to the four kinetic equations, which requires, unfortunately, cumbersome calculations, is the main topic of Sec. 3. If the reader is not interested in the discussion of the approximations employed in this analysis and the derivation of basic equations (26)–(29), which will be used in Secs. 6 and 7, he can omit Sec. 3 and skip to Sec. 4, which analyzes spontaneous and light-induced tunneling in TLS on the basis of the four kinetic equations taking into account only one TLS and shows that the light-induced tunneling can be neglected at low intensities of pumping. This conclusion makes considerably easier the development of an approximate method for describing a molecule's interaction with a larger number of TLSs in the subsequent sections. In Sec. 5, an approximate effective technique will be developed for determining the kinetics of a system consisting of one molecule and one TLS without light-induced tunneling. In Sec. 6 this technique will be generalized to the case of interaction between one molecular chromophore and many TLSs, which will allow us to obtain an expression for the light absorption coefficient with due account of the time-dependent broadening of spectral lines, i.e., the spectral diffusion due to interaction with a lot of nonequilibrium TLSs. On the basis of the equations derived in Secs. 3 and 6, an expression for the two-photon correlator will be derived in Sec. 7 taking into account the electron-phonon coupling, TLS relaxation, and effect of spectral diffusion. Two specific examples of its applications will be given there. Section 8 summarizes the paper and presents conclusions.

2. HAMILTONIAN OF THE ELECTRON-PHONON-TUNNELON SYSTEM

Let us consider a system consisting of a two-level chromophore interacting with a transverse electromagnetic field, vibrations of nuclei in the lattice (phonons), and tunneling transitions in the solvent (tunnelons). The Hamiltonian of such a system has the form

$$H=H_0+H_\perp+\hat{\Lambda}+\hat{\lambda}, \tag{2}$$

where $H_\perp$ is the Hamiltonian of the transverse electromagnetic field and H_0 is the Hamiltonian of the chromophore-phonon-tunnelon system:

$$H_0=[\hbar\Omega+\Delta H(\xi)]B^+B+H(\xi). \tag{3}$$

Here Ω is the frequency at which the electronic system of the chromophore is excited, B^+ and B are creation and annihilation operators of the electronic excitation, $H(\xi)$ is the Hamiltonian of phonons and TLS. The function $\Delta H(\xi)$ describes changes in the phonon-tunnelon system when the chromophore's electronic system is excited. This interaction of the Franck–Condon type is determined by changes in the adiabatic potentials when the electronic system is excited. The operator $\hat{\Lambda}=\mathbf{d}\cdot\mathbf{E}(B+B^+)$ describes an interaction between the optically active electron in the chromophore and light, and $\hat{\lambda}=\lambda(c+c^+)$ is the operator of tunneling in TLS.

The Hamiltonian of the phonon-tunnelon system has the form

$$H(\xi)=[\hbar\varepsilon+V(R)]c^+c+H(R). \tag{4}$$

Here $\hbar\varepsilon$ is the tunnelon energy, c^+ and c are the creation and annihilation operators of excitations in TLS, i.e. tunnelons, $H(R)$ is the phonon Hamiltonian, and $V(R)$ characterizes changes in the phonon Hamiltonian due to excitations in TLS, i.e., it is an interaction of the Franck–Condon type in

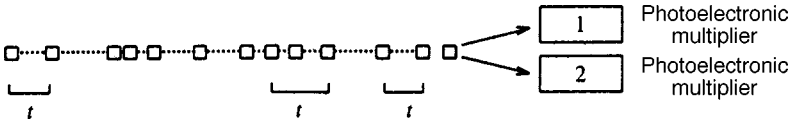

FIG. 2. Train of photons emitted by a single molecule excited by a cw laser. Pairs of photons with the same delay are shown.

the ground state of the chromophore electronic system which is diagonal with respect to the tunnelon operators.

Strictly speaking, the tunneling operator $\hat{\lambda}=\lambda(c+c^+)$ of the tunnelon-phonon system, which is represented by a separate term on the right of Eq. (2) expressing the full Hamiltonian, should also be included in the phonon-tunnelon Hamiltonian. There is a good reason, however, to express it by a separate term in the full Hamiltonian because the latter is used in deriving the full set of equations for the density operator, so one can easily see what new terms are generated in the equations by this tunneling operator.

As was noted above, the adiabatic potential of the chromophore-phonon-tunnelon system changes when the chromophore's electronic system is excited, therefore, an additional term turns up in the tunnelon-phonon Hamiltonian $H(\xi)$:

$$\Delta H(\xi)=[\hbar\Delta+\Delta V(R)]c^+c+\Delta H(R). \tag{5}$$

Here $\hbar\Delta$ is the change in the splitting in TLS, i.e., the change in the tunnelon system. This parameter characterizes the interaction, which is quadratic in the electron and tunnelon operators. The term $\Delta H(R)$ is the operator of electron-phonon interaction, and $\Delta V(R)$ determines changes in the tunnelon-phonon interaction due to the electronic excitation. The latter contribution is required because the rates of tunneling transitions in the ground and electronically excited states are different.

The last two terms on the right-hand side of Eq. (2) characterize, as was noted above, the interaction between the chromophore and transverse electromagnetic field and the tunneling operator. Indeed, if we set $\hat{\Lambda}=\hat{\lambda}=0$, transitions between two chromophore states and in TLS become impossible.

On the other hand, if we set $\hat{\Lambda}\neq 0$ but $\hat{\lambda}=0$, transitions in the electronic system become possible, but tunneling transitions in TLSs are not allowed. This is the case for which the density operator was calculated previously[6] without taking account of TLS. Now we are facing the problem of deriving equations for the density operator with due account of TLSs and tunneling operator $\hat{\lambda}=\lambda(c+c^+)$. This problem will be solved in the next section.

3. EQUATIONS FOR THE DENSITY MATRIX OF AN ELECTRON-PHONON-TUNNELON SYSTEM

Now consider the equation for the density matrix of the entire system:

$$i\hbar\dot{\hat{\rho}}=[H,\hat{\rho}], \tag{6}$$

where H is given by Eq. (2). In order to derive equations for the matrix elements from this basic equation, we have to choose a definite basis. Let us choose the eigenfunctions of the operator $H_0+H_\perp$, which does not account for tunneling and electronic transitions, for such a basis. These functions are products of electron, phonon, and tunnelon wave functions, as well as oscillatory functions describing photons.

The electron functions of a two-level chromophore are determined by the equations

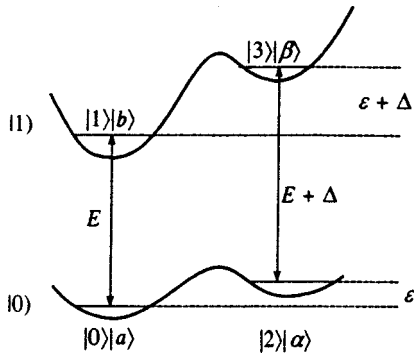

FIG. 3. Electron-phonon-tunnelon wave functions and corresponding energy levels.

$$B^+|0)=|1), \quad B^+|1)=0,$$
$$B|1)=0, \quad B|1)=|0). \tag{7}$$

The phonon functions in the ground electronic state are derived from the following two equations corresponding to different states of TLS:

$$H(R)|a\rangle=\hbar\Omega_a|a\rangle,$$
$$[H(R)+V(R)]|\alpha\rangle=\hbar\Omega_\alpha|\alpha\rangle, \tag{8}$$

and in the excited electronic state from the other two equations:

$$[H(R)+\Delta H(R)]|b\rangle=\hbar\Omega_b|b\rangle,$$
$$[H(R)+V(R)+\Delta V(R)+\Delta H(R)]|\beta\rangle=\hbar\Omega_\beta|\beta\rangle. \tag{9}$$

The tunnelon wave functions in the ground and excited electronic states satisfy the equations

$$\varepsilon c^+c|0\rangle=0, \quad (\varepsilon+\Delta)c^+c|1\rangle=0,$$
$$\varepsilon c^+c|2\rangle=\varepsilon|2\rangle, \quad (\varepsilon+\Delta)c^+c|3\rangle=(\varepsilon+\Delta)|3\rangle. \tag{10}$$

The eigenfunctions $|\mathbf{n}\rangle=|n_1,n_2,\ldots\rangle$ of the Hamiltonian $H_\perp$ of transverse electromagnetic field are the harmonic oscillator functions.

It is obvious that the eigenfunctions of Hamiltonian $H_0+H_\perp$ have the form

$$|A)=|\mathbf{n}\rangle|0)|A\rangle=|\mathbf{n}\rangle|0)\begin{cases}|0\rangle|a\rangle,\\ |2\rangle|\alpha\rangle,\end{cases}$$
$$|B)=|\mathbf{n}-1\rangle|1)|B\rangle=|\mathbf{n}-1\rangle|1)\begin{cases}|1\rangle|b\rangle,\\ |3\rangle|\beta\rangle,\end{cases} \tag{11}$$

where the wave functions of the system containing the chromophore, tunnelons, and phonons satisfy the equations

$$H_0|0)|A\rangle=\hbar\Omega_A|0)|A\rangle,$$
$$H_0|1)|B\rangle=(E+\hbar\Omega_B)|1)|B\rangle, \tag{12}$$

where

$$\Omega_A=\begin{cases}\Omega_a,\\ \Omega_\alpha+\varepsilon,\end{cases} \quad \Omega_B=\begin{cases}\Omega_b,\\ \Omega_\beta+\varepsilon+\Delta.\end{cases} \tag{13}$$

This set of functions and a diagram of corresponding energy levels are shown in Fig. 3. Using the earlier investigations,[5,6] where spontaneous light emission was taken into account in

deriving equations for the density operator, and basis (11), we transform operator equation (6) to the following set of equations:

$$\dot\rho_{BA}=-i\left(\Delta_0+\Omega_{BA}-\frac{i}{2T_1}\right)\rho_{BA}$$
$$-i\sum_{A'}\Lambda_{BA'}\rho_{A'A}+i\sum_{B'}\rho_{BB'}\Lambda_{B'A}$$
$$-i\sum_{B'}\lambda_{BB'}\rho_{B'A}+i\sum_{A'}\rho_{BA'}\lambda_{A'A},$$

$$\dot\rho_{AB}=-i\left(-\Delta_0+\Omega_{AB}-\frac{i}{2T_1}\right)\rho_{AB}$$
$$-i\sum_{B'}\Lambda_{AB'}\rho_{B'B}+i\sum_{A'}\rho_{AA'}\Lambda_{A'B}$$
$$-i\sum_{A'}\lambda_{AA'}\rho_{A'B}+i\sum_{B'}\rho_{AB'}\lambda_{B'B},$$

$$\dot\rho_{BB'}=-i\left(\Omega_{BB'}-\frac{i}{T_1}\right)\rho_{BB'}-i\sum_{A}(\Lambda_{BA}\rho_{AB'}-\rho_{BA}\Lambda_{AB'})$$
$$-i\sum_{B''}(\lambda_{BB''}\rho_{B''B'}-\rho_{BB''}\lambda_{B''B'}),$$

$$\dot\rho_{AA'}=-i\Omega_{AA'}\rho_{AA'}+\frac{1}{T_1}\sum_{BB'}\langle A|B\rangle\rho_{BB'}\langle B'|A'\rangle$$
$$-i\sum_{B}(\lambda_{AB}\rho_{BA}-\rho_{AB}\lambda_{BA'})$$
$$-i\sum_{A'}(\lambda_{AA''}\rho_{A''A'}-\rho_{AA''}\lambda_{A''A'}). \quad (14)$$

Here $\Delta_0=\Omega-\omega_0$ is the difference between the electronic resonant frequency and absorbed photon frequency. The matrix elements and frequencies in this set of equations are given by the following relations:

$$\Lambda_{BA}=\langle B|A\rangle\chi=\begin{cases}\langle b|a\rangle\langle 1|0\rangle\chi=\Lambda_{ba},\\ \langle\beta|\alpha\rangle\langle 3|2\rangle\chi=\Lambda_{\beta\alpha},\end{cases}\quad \Lambda_{AB}=\Lambda_{BA}^*, \quad (15)$$

$$\lambda_{BB'}=\begin{cases}\langle b|\beta\rangle\langle 1|3\rangle\lambda=\lambda_{b\beta},\\ \langle\beta|b\rangle\langle 3|1\rangle\lambda=\lambda_{\beta b},\end{cases}$$

$$\lambda_{AA'}=\begin{cases}\langle a|\alpha\rangle\langle 0|2\rangle\lambda=\lambda_{a\alpha},\\ \langle\alpha|a\rangle\langle 2|0\rangle\lambda=\lambda_{\alpha a},\end{cases} \quad (16)$$

$$\Omega_{BA}=\Omega_B-\Omega_A,\quad \Omega_{BB'}=\Omega_B-\Omega_{B'},\quad \Omega_{AA'}=\Omega_A-\Omega_{A'}. \quad (17)$$

Here $\chi=\mathbf{d}\cdot\mathbf{E}/\hbar$ is the Rabi frequency. The underlined components of Eq. (12) are generated by the tunneling operator. If these terms are omitted, the resulting set of equations is essentially identical to the set of equations (19) in Ref. 6.

Now let us write down each of the four lines of Eq. (12) in a more detailed form. To this end, we introduce for simplicity the following notation: $\rho_{b1a0}=\rho_{ba}$, $\rho_{\beta3\alpha2}=\rho_{\beta\alpha}$, $\rho_{b1\beta3}=\rho_{b\beta}$, $\rho_{a0\alpha2}=\rho_{a\alpha}$,... for the matrix elements of the density matrix and two approximations. First, we take into account only vertical optical transitions shown in Fig. 3. These optical transitions are not assisted by simultaneous tunneling processes. Then all matrix elements and density matrix elements can be equated to zero: $\Lambda_{a\beta}=\Lambda_{b\alpha}=\rho_{a\beta}=\rho_{b\alpha}=0$. Second, we also omit the off-diagonal matrix elements of the density matrix, $\rho_{aa'}$, $\rho_{bb'}$, $\rho_{\alpha\alpha'}$, and $\rho_{\beta\beta'}$, which are inessential for the analysis of the effect of operators $\hat\Lambda$ and $\hat\lambda$ on the diagonal elements of the density matrix in the lowest nonvanishing order, and this approximation will be used hereafter.

With due account of these approximations, let us proceed to a more detailed form of the four equations in (14). From the first and second equations, we obtain

$$\dot\rho_{ba}=-i(\Delta_0+\Omega_{ba}-i/2T_1)\rho_{ba}-i\Lambda_{ba}(\rho_{aa}-\rho_{bb}),$$
$$\dot\rho_{\beta\alpha}=-i(\Delta_0+\Delta+\Omega_{\beta\alpha}-i/2T_1)\rho_{\beta\alpha}-i\Lambda_{\beta\alpha}(\rho_{\alpha\alpha}-\rho_{\beta\beta}),$$
$$\dot\rho_{ab}=-i(-\Delta_0+\Omega_{ab}-i/2T_1)\rho_{ab}-i\Lambda_{ab}(\rho_{bb}-\rho_{aa}),$$
$$\dot\rho_{\alpha\beta}=-i(-\Delta_0-\Delta+\Omega_{\alpha\beta}-i/2T_1)\rho_{\alpha\beta}-i\Lambda_{\alpha\beta}(\rho_{\beta\beta}-\rho_{\alpha\alpha}). \quad (18)$$

The third equation transforms to the following four equations:

$$\dot\rho_{b\beta}=-i(-\varepsilon-\Delta+\Omega_{b\beta}-i0)\rho_{b\beta}-i\lambda_{b\beta}(\rho_{\beta\beta}-\rho_{bb}),$$
$$\dot\rho_{\beta b}=-i(\varepsilon+\Delta+\Omega_{\beta b}-i0)\rho_{\beta b}-i\lambda_{\beta b}(\rho_{bb}-\rho_{\beta\beta}),$$
$$\dot\rho_{bb}=-\frac{\rho_{bb}}{T_1}-i\sum_{a}(\Lambda_{ba}\rho_{ab}-\rho_{ba}\Lambda_{ab})$$
$$-i\sum_{\beta}(\lambda_{b\beta}\rho_{\beta b}-\rho_{b\beta}\lambda_{\beta b}),$$
$$\dot\rho_{\beta\beta}=-\frac{\rho_{\beta\beta}}{T_1}-i\sum_{\alpha}(\Lambda_{\beta\alpha}\rho_{\alpha\beta}-\rho_{\beta\alpha}\Lambda_{\alpha\beta})$$
$$-i\sum_{b}(\lambda_{\beta b}\rho_{b\beta}-\rho_{\beta b}\lambda_{b\beta}). \quad (19)$$

Finally, the fourth equation in (14) transforms to the following four equations:

$$\dot\rho_{a\alpha}=-i(-\varepsilon+\Omega_{a\alpha}-i0)\rho_{a\alpha}-i\lambda_{a\alpha}(\rho_{\alpha\alpha}-\rho_{aa}),$$
$$\dot\rho_{\alpha a}=-i(\varepsilon+\Omega_{\alpha a}-i0)\rho_{\alpha a}-i\lambda_{\alpha a}(\rho_{aa}-\rho_{\alpha\alpha}),$$
$$\dot\rho_{aa}=\frac{1}{T_1}\sum_{b}\langle a|b\rangle\langle 0|1\rangle\rho_{bb}\langle 1|0\rangle\langle b|a\rangle-i\sum_{b}(\Lambda_{ab}\rho_{ba}$$
$$-\rho_{ab}\Lambda_{ba})-i\sum_{\alpha}(\lambda_{a\alpha}\rho_{\alpha a}-\rho_{a\alpha}\lambda_{\alpha a}),$$
$$\dot\rho_{\alpha\alpha}=\frac{1}{T_1}\sum_{\beta}\langle\alpha|\beta\rangle\langle 2|3\rangle\rho_{\beta\beta}\langle 3|2\rangle\langle\beta|\alpha\rangle$$
$$-i\sum_{\beta}(\Lambda_{\alpha\beta}\rho_{\beta\alpha}-\rho_{\alpha\beta}\Lambda_{\beta\alpha})$$
$$-i\sum_{a}(\lambda_{\alpha a}\rho_{a\alpha}-\rho_{\alpha a}\lambda_{a\alpha}). \quad (20)$$

Here and in the previous equations, the underlined terms are those generated by the tunneling operator.

Under low-intensity pumping, the kinetics of population numbers can be investigated neglecting coherence effects, which is equivalent to omitting time derivatives of the off-diagonal elements of the density matrix. In the case under consideration, both the pumping intensity and tunneling operator can be treated as small quantities. Therefore, a third approximation can be applied to Eq. (18) and underlined terms of Eqs. (19) and (20), specifically, we can set $\dot{\rho}_{ba}=\dot{\rho}_{ab}=\dot{\rho}_{\beta\alpha}=\dot{\rho}_{\alpha\beta}=\dot{\rho}_{b\beta}=\dot{\rho}_{\beta b}=\dot{\rho}_{a\alpha}=\dot{\rho}_{\alpha a}=0$. After that, one can easily derive from these equations expressions for off-diagonal matrix elements and substitute them in the last two pairs of equations (19) and (20). Then we have the following set of rate equations:

$$\dot{\rho}_{bb}=-\left(\frac{1}{T_1}+\sum_a k_{ba}\right)\rho_{bb}+\sum_a k_{ba}\rho_{aa}$$
$$-\sum_\beta r_{b\beta}\rho_{bb}+\sum_\beta r_{b\beta}\rho_{\beta\beta},$$
$$\dot{\rho}_{aa}=\frac{1}{T_1}\sum_b \langle a|b\rangle\langle 0|1\rangle\rho_{bb}\langle 1|0\rangle\langle b|a\rangle+\sum_b k_{ba}\rho_{bb}$$
$$-\sum_b k_{ba}\rho_{aa}-\sum_\alpha r_{a\alpha}\rho_{aa}+\sum_\alpha r_{a\alpha}\rho_{\alpha\alpha},$$
$$\dot{\rho}_{\beta\beta}=-\left(\frac{1}{T_1}+\sum_\alpha k_{\beta\alpha}\right)\rho_{\beta\beta}+\sum_\alpha k_{\beta\alpha}\rho_{\alpha\alpha}$$
$$+\sum_b r_{b\beta}\rho_{bb}-\sum_b r_{b\beta}\rho_{\beta\beta},$$
$$\dot{\rho}_{\alpha\alpha}=\frac{1}{T_1}\sum_\beta \langle\alpha|\beta\rangle\langle 2|3\rangle\rho_{\beta\beta}\langle 3|2\rangle\langle\beta|\alpha\rangle+\sum_\beta k_{\beta\alpha}\rho_{\beta\beta}$$
$$-\sum_\beta k_{\beta\alpha}\rho_{\alpha\alpha}+\sum_a r_{a\alpha}\rho_{aa}-\sum_a r_{a\alpha}\rho_{\alpha\alpha}. \qquad (21)$$

Here

$$k_{ba}=\Lambda_{ba}\Lambda_{ab}\frac{1/T_1}{(\Delta_0+\Omega_{ba})^2+(1/2T_1)^2},$$
$$k_{\beta\alpha}=\Lambda_{\beta\alpha}\Lambda_{\alpha\beta}\frac{1/T_1}{(\Delta_0+\Delta+\Omega_{\beta\alpha})^2+(1/2T_1)^2} \qquad (22)$$

characterize the rates of optical transitions, and

$$r_{b\beta}=2\pi\lambda_{b\beta}\lambda_{\beta b}\delta(\Omega_{b\beta}-\varepsilon-\Delta),$$
$$r_{a\alpha}=2\pi\lambda_{a\alpha}\lambda_{\alpha a}\delta(\Omega_{a\alpha}-\varepsilon) \qquad (23)$$

are the rates of tunneling transitions.

It is obvious that we can express the diagonal elements of the density matrix in (21) in the form

$$\rho_{aa}=\rho_a\rho_0,\quad \rho_{bb}=\rho_b\rho_1,\quad \rho_{\alpha\alpha}=\rho_\alpha\rho_2,\quad \rho_{\beta\beta}=\rho_\beta\rho_3, \qquad (24)$$

where ρ_a, ρ_b, ρ_α, and ρ_β are the probabilities of finding the system in the corresponding phonon states. They satisfy the condition

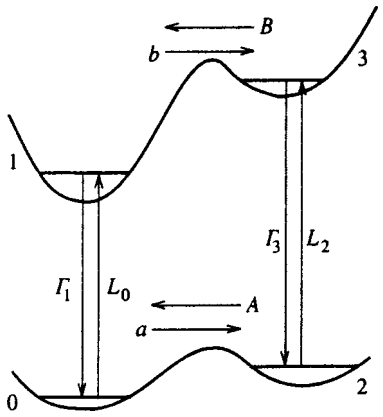

FIG. 4. Optical and tunneling transitions described by Eqs. (26).

$$\sum_a \rho_a=\sum_b \rho_b=\sum_\alpha \rho_\alpha=\sum_\beta \rho_\beta=1. \qquad (25)$$

By substituting (24) in the set of equations (21) and summing over the phonon indices, we obtain the following rate equations:

$$\dot{\rho}_1=-\Gamma_1\rho_1+L_0\rho_0-b\rho_1+B\rho_3,$$
$$\dot{\rho}_0=\Gamma_1\rho_1-L_0\rho_0-\alpha\rho_0+A\rho_2,$$
$$\dot{\rho}_3=-\Gamma_3\rho_3+L_2\rho_2+b\rho_1-B\rho_3,$$
$$\dot{\rho}_2=\Gamma_3\rho_3-L_2\rho_2+a\rho_0-A\rho_2, \qquad (26)$$

where

$$\Gamma_1=1/T_1+L_1,\quad \Gamma_3=1/T_1+L_3, \qquad (27)$$

and the coefficients

$$L_0=k_{10}^g=\sum_{b,a}\rho_a k_{ba}$$
$$=\langle 1|0\rangle^2\chi^2\sum_{b,a}\rho_a\langle b|a\rangle^2\frac{1/T_1}{(\Delta_0+\Omega_{ba})^2+(1/2T_1)^2},$$
$$L_1=k_{10}^e=\sum_{ba}\rho_b k_{ba},$$
$$L_2=k_{32}^g=\sum_{\beta,\alpha}\rho_\alpha k_{\beta\alpha}$$
$$=\langle 3|2\rangle^2\chi^2\sum_{\beta,\alpha}\rho_\alpha\langle\beta|\alpha\rangle^2\frac{1/T_1}{(\Delta_0+\Delta+\Omega_{\beta\alpha})^2+(1/2T_1)^2},$$
$$L_3=k_{32}^e=\sum_{\beta,\alpha}\rho_\beta k_{\beta\alpha} \qquad (28)$$

characterize the rates of induced optical transitions between the pairs of levels (1,0) and (3,2), as shown in Fig. 4. The coefficients

$$A=\sum_{\alpha a}\rho_\alpha r_{a\alpha}=2\pi\sum_{\alpha a}\rho_\alpha\lambda_{a\alpha}\lambda_{\alpha a}\delta(\Omega_{a\alpha}-\varepsilon),$$
$$a=\sum_{\alpha a}\rho_a r_{a\alpha},$$

$$B=\sum_{\beta b}\rho_\beta r_{b\beta}=2\pi\sum_{\beta b}\rho_\beta\lambda_{b\beta}\lambda_{\beta b}\delta(\Omega_{b\beta}-\varepsilon-\Delta),$$

$$b=\sum_{\beta b}\rho_b r_{b\beta} \tag{29}$$

are the desired microscopic expressions for the rates of tunneling transitions between the levels in the upper and lower potential wells, respectively, specifically, the rates b and B refer to tunneling transitions in the excited electronic state and the rates a and A to the transitions in the ground state. The rates of optical and tunneling transitions determined by Eqs. (27)–(29) are indicated in Fig. 4 next to the arrows corresponding to these transitions. The letters A and B denote the rates of tunneling transitions from higher to lower potential wells. These transition rates do not vanish at zero temperature and are higher than the rates a and b of transitions from lower to higher wells, which tend to zero at zero temperature.

4. SPONTANEOUS AND LIGHT-INDUCED TRANSITIONS IN A TWO-LEVEL SYSTEM

Generally speaking, the set of kinetic equations (26), which is the main result of the previous section, could be written without complicated calculations using the diagram of transitions shown in Fig. 4. Using this approach, however, we would not have obtained (a) microscopic expressions for the constants in these kinetic equations, which will be used in what follows, and (b) a clear understanding of what approximations underlie Eqs. (26). The latter can be used as a starting point for further analysis.

Rate equations (26) take into account the tunneling in both ground and electron-excited chromophore states. Let us estimate the relative contributions of these processes. Consider the case in which the system is excited only at the frequency of transition 1–0, i.e., $L_2=L_3=0$. Let us assume also that level 3 in Fig. 4 is lower than the excited level 1. The latter condition is necessary so that the model could be applied to the situation with jumps of the spectral line like those shown in Fig. 1. After such a modification of the model, the constants B and b should be interchanged, and rate equations (26) transform to

$$\begin{aligned}
\dot\rho_1&=-(\Gamma_1+B)\rho_1+L_0\rho_0+b\rho_3,\\
\dot\rho_3&=B\rho_1-(1/T_1+b)\rho_3,\\
\dot\rho_0&=\Gamma_1\rho_1-(L_0+a)\rho_0+A\rho_2 a,\\
\dot\rho_2&=a\rho_0+\rho_3/T_1-A\rho_2.
\end{aligned} \tag{30}$$

Let us assume that the transition rate constants satisfy the inequalities

$$\Gamma\gg L_0\gg A,\quad B>a,b, \tag{31}$$

which are quite common for real systems. If the rate constants satisfy these conditions, the probability ρ_j changes varies time on two scales: the shorter and longer ones, on which the relaxation rate of ρ_j is faster and slower, respectively. The fast relaxation occurs on a time scale of the order of T_1. After that, quasi-equilibrium is established between the populations ρ_1 and ρ_3 of electron-excited states on the one hand, and the population ρ_0 on the other. The required relation can be obtained by setting $\dot\rho_1=\dot\rho_3=0$. Then, taking into account inequalities (31), we derive from the first and second equations in (30)

$$\begin{aligned}
\rho_1&=\frac{(1/T_1+b)L_0\rho_0}{(\Gamma_1+B)/T_1+b\Gamma_1}\simeq\frac{L_0}{\Gamma_1}\rho_0,\\
\rho_3&=\frac{BL_0\rho_0}{(\Gamma_1+B)/T_1+b\Gamma_1}\simeq\frac{L_0}{\Gamma_1}T_1B\rho_0.
\end{aligned} \tag{32}$$

After substituting these results in the remaining two equations in (30), we obtain the equations

$$\begin{aligned}
\dot\rho_0&=-(\tilde B+a)\rho_0+A\rho_2,\\
\dot\rho_2&=(\tilde B+a)\rho_0-A\rho_2,
\end{aligned} \tag{33}$$

whose solution is

$$\begin{aligned}
\rho_0(t)&=\frac{A}{\tilde B+R}+\left[\rho_0(0)-\frac{A}{\tilde B+R}\right]\exp[-(\tilde B+R)t],\\
\rho_2(t)&=1-\rho_0(t).
\end{aligned} \tag{34}$$

Here

$$\tilde B=\frac{L_0}{\Gamma_1}B,\quad R=A+a. \tag{35}$$

According to Eqs. (32) and (34), the slow relaxation rate of all populations is determined both by tunneling between states 0 and 2 and between 1 and 3. The transition between states 1 and 3 is light-induced tunneling. Its rate $\tilde B$ is proportional to the pumping L_0.

Now consider the case in which the pumping frequency resonates with the 3–2 transition. Then rate equations (26) take the form

$$\begin{aligned}
\dot\rho_1&=-(1/T_1+B)\rho_1+b_3,\\
\dot\rho_3&=B\rho_1-(\Gamma_3+b)\rho_3+L_2\rho_2,\\
\dot\rho_0&=\rho_1/T_1-a\rho_0+A\rho_2,\\
\dot\rho_2&=\Gamma_3\rho_3+a\rho_0-(L_2+A)\rho_2.
\end{aligned} \tag{36}$$

Let us solve these equations similarly to Eqs. (30). After setting $\dot\rho_1=\dot\rho_3=0$, we derive from (36)

$$\begin{aligned}
\rho_1&=\frac{bL_2\rho_2}{(\Gamma_3+b)/T_1+B\Gamma_3}\simeq\frac{L_2}{\Gamma_3}T_1b\rho_2,\\
\rho_3&=\frac{(1/T_1+B)L_2\rho_2}{(\Gamma_3+b)/T_1+B\Gamma_3}\simeq\frac{L_2}{\Gamma_3}\rho_2.
\end{aligned} \tag{37}$$

With due account of these results, the last two equations in (36) transform to

$$\begin{aligned}
\dot\rho_0&=-a\rho_0+(\tilde b+A)\rho_2,\\
\dot\rho_2&=a\rho_0-(\tilde b+A)\rho_2,
\end{aligned} \tag{38}$$

where

$$\tilde{b}=\frac{L_2}{\Gamma_3}b \tag{39}$$

accounts for the contribution of light-induced transitions to the tunneling when the pumping light frequency coincides with that of the 3–2 transition. The solution of Eqs. (38) is

$$\rho_2(t+t_0)=\frac{a}{\tilde{b}+R}+\left[\rho_2(t_0)-\frac{a}{\tilde{b}+R}\right]\exp[-(\tilde{b}+R)t],$$

$$\rho_0(t+t_0)=1-\rho_2(t+t_0). \tag{40}$$

With a view to investigate later the situation illustrated by Fig. 1, we set the initial moment of time not to zero, but t_0. We describe in Sec. 7 how one can account for spectral line jumps like those shown in Fig. 1 with the help of Eqs. (32) and (34), and with Eqs. (37) and (40) as well.

The rates of light-induced transitions can be higher than those of spontaneous ones. This is the case of TLSs involved in photochemical burning of holes in inhomogeneously broadened optical bands. Such TLSs usually describe the states of chromophores themselves, i.e., these are TLSs due to embedding of impurity molecules in a solvent (extrinsic two-level systems). Obviously, the number of such extrinsic TLSs approximately equals the number of impurities. In this situation, as follows from experimental data, a steady-state spectral hole is burnt in an inhomogeneously broadened optical band owing to the light-induced tunneling. Since such holes persist at low temperatures and for several days and even weeks after the optical pumping is turned off, this means that the ''dark'' tunneling in the ground electronic state has a very low efficiency, i.e., the constants a and A are low in such extrinsic TLSs. Consequently, the light-induced tunneling cannot be neglected in such TLSs.

In polymers and glasses, however, there are TLSs of a different type, which are inherent to a solvent (and called intrinsic two-level systems) and whose existence is not caused by the presence of impurities. The number of such TLSs can be enormous and much larger than that of impurities. If an isolated TLS discussed in this section is of such a nature, the spontaneous tunneling rate constants b and B in this TLS in the excited state of a chromophore are comparable to the tunneling constants a and A in the ground state. Then, in accordance with Eqs. (35) and (39), one can neglect light-induced transitions between states 1 and 3 since $L/\Gamma \ll 1$. This situation takes place in a chromophore coupled to TLSs in a polymer or glass. How to take account of the chromophore's interaction with the set of nonequilibrium TLSs inherent to polymers then becomes a problem, which will be discussed in the next two sections.

5. INTERACTION WITH ONE TWO-LEVEL SYSTEM

Before proceeding to solving one of the central problems of the reported investigation, namely, the problem of taking into account interaction with the great number of nonequilibrium TLSs, let us radically simplify rate equations (30), which take account of only one TLS. After neglecting the tunneling transition in the excited electronic state, which is possible in the case of an intrinsic TLS, we derive from Eq. (30) the following simplified equations:

$$\dot{\rho}_1=-\Gamma_1\rho_1+L_0\rho_0,$$

$$\dot{\rho}_0=\Gamma_1\rho_1-L_0\rho_0-a\rho_0+A\rho_2,$$

$$\dot{\rho}_2=a\rho_0-A\rho_2, \tag{41}$$

whose exact (but complicated) solution can be found. With due account of conditions (31) and in the zero-order approximation in the small parameter R/Γ, this exact solution can be transformed to the following form:

$$\rho_1(t)=n_1(t)-n_1(\infty)p_2(t),$$

$$\rho_2(t)=n_0(\infty)p_2(t),$$

$$\rho_0(t)=1-\rho_1(t)-\rho_2(t)$$

$$=n_0(t)-n_0(\infty)p_2(t)+n_1(\infty)p_2(t), \tag{42}$$

where the functions

$$n_1(t)=\frac{L_0}{\Gamma+L_0}\{1-\exp[-(\Gamma_1+L_0)t]\},$$

$$n_0(t)=1-n_1(t) \tag{43}$$

describe the time dependence of the probabilities controlled by laser pumping, but with the tunneling mechanism ''turned off,'' and the functions

$$p_2(t)=\frac{a}{R}(1-e^{-Rt}),\quad p_0(t)=1-p_2(t), \tag{44}$$

on the contrary, determine the evolution of the probabilities controlled by tunneling, but with the optical pumping ''turned off.'' This separation of the relaxation processes of the electronic and tunneling degrees of freedom in the approximate formulas (42) will allow us to obtain a relatively simple generalization to the case of many extrinsic TLSs.

At $p_2(0)=0$, the function $p_2(t)$ rises on a time scale of order Γ^{-1} only to values of the order of $R/\Gamma \ll 1$. Therefore we can replace some variables in Eqs. (42):

$$n_0(\infty)p_2(t)\to n_0(t)p_2(t),\quad n_1(\infty)p_2(t)\to n_1(t)p_2(t). \tag{45}$$

Then Eqs. (42) transform to

$$\rho_1(t)=n_1(t)p_0(t),\quad \rho_2(t)=n_0(t)p_2(t),$$

$$\rho_0(t)=n_1(t)[1-p_0(t)]+n_0(t)p_0(t). \tag{46}$$

The difference between these formulas and (42), derived from the exact solution of Eqs. (41), is of the order of the very small parameter R/Γ. The solution (46) is more convenient than the exact solution of Eqs. (41) because the former is based on the functions $n_{1,0}(t)$ and $p_{0,2}(t)$, which were obtained under the condition that the tunneling and laser pumping do not influence one another. This property of Eqs. (46) allows us to draw up a prescription for finding an approximate solution in the case in which a chromophore interacts with a set of TLSs and calculation of the exact solution is very difficult.

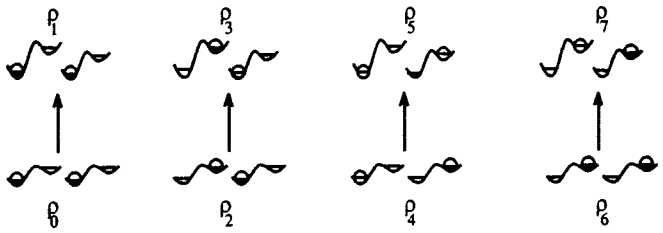

FIG. 5. Possible quantum states of a system of two TLSs and allowed direct optical transitions without tunneling.

The underlying idea of the approximate solution is that the equations (30) consist of two almost decoupled blocks: the ''electronic'' block with large relaxation rate constants Γ and L, and the ''tunneling'' block with relatively small relaxation rate constants A and a.

a) First, neglecting the tunneling constants in Eqs. (41), we solve the equations with the electronic rate constants:

$$\dot{\rho}_1 = -\Gamma_1\rho_1 + L_0\rho_0, \quad \dot{\rho}_0 = \Gamma_1\rho_1 - L_0\rho_0. \tag{47}$$

The solution is hereafter denoted by $n_{1,0}(t)$. This solution of the electronic block adequately describes the evolution of exact probabilities at small times $t \sim \Gamma^{-1}$.

b) Then, neglecting the optical pumping and electronic relaxation rate constants, we derive from Eq. (41) the equations

$$\dot{\rho}_0 = -a\rho_0 + A\rho_2, \quad \dot{\rho}_2 = a\rho_0 - A\rho_2, \tag{48}$$

whose solution is denoted by $p_{0,2}(t)$. This solution of the tunneling block describes the system evolution at large times $t \sim R^{-1}$.

c) Finally, as was shown above, one can construct an approximate solution from the combination of Eqs. (47) and (48) with due account of Eqs. (46), and this solution is fairly close to the exact solution of Eqs. (41) over the entire time interval.

6. INTERACTION WITH A MULTITUDE OF TWO-LEVEL SYSTEMS

Let us first apply the suggested prescription to calculation of approximate solution of a system consisting of one chromophore interacting with two TLSs. Since one TLS has two quantum states, the system of two TLSs has four states. Figure 5 shows these four states of two TLSs in both ground and excited states of the chromophore. Let us label these states as shown in Fig. 5.

In Sec. 3 we replaced the infinite set of equations (14) with Eqs. (26), which contain only four lines; concurrently we derived formulas for the relaxation rate constants. A similar procedure can be performed in the case of a chromophore interacting with two TLSs. But now we have, instead of four equations, a system of eight equations because the number of the states of two TLSs in combination with two electronic states is, according to Fig. 5, eight.

As in the previous sections, we limit our analysis to vertical optical transitions, i.e., those not assisted by tunneling transitions. These vertical transitions include only those shown in Fig. 5. All other optical transitions are assisted by a tunneling transition in one TLS, i.e., their rates are very low.

The states of TLS in a given electronic state can be coupled only by the tunneling operator $\hat{\lambda} = \lambda(c + c^+)$. Since this operator is linear in the tunnelon creation and annihilation operators, it can only couple states that differ by one tunnelon, i.e.,

$$\underbrace{\rho_0 - \rho_2 - \rho_6 - \rho_4}, \qquad \underbrace{\rho_1 - \rho_3 - \rho_7 - \rho_5}.$$

These comments explain why the system of a chromophore and two TLSs is described by the following eight equations:

$$\begin{aligned}
\dot{\rho}_1 &= -(\Gamma_1 + b + b')\rho_1 + L_0\rho_0 + B\rho_3 + B'\rho_5,\\
\dot{\rho}_0 &= \Gamma_1\rho_1 - (L_0 + a + a')\rho_0 + A\rho_2 + A'\rho_4,\\
\dot{\rho}_3 &= -(\Gamma_3 + B + b')\rho_3 + L_2\rho_2 + b\rho_1 + B'\rho_7,\\
\dot{\rho}_2 &= \Gamma_3\rho_3 - (L_2 + A + a')\rho_2 + a\rho_0 + A'\rho_6,\\
\dot{\rho}_5 &= -(\Gamma_5 + B' + b)\rho_5 + L_4\rho_4 + b'\rho_1 + B\rho_7,\\
\dot{\rho}_4 &= \Gamma_5\rho_5 - (L_4 + A' + a)\rho_4 + a'\rho_0 + A\rho_6,\\
\dot{\rho}_7 &= -(\Gamma_7 + B' + B)\rho_7 + L_6\rho_6 + b'\rho_3 + b\rho_5,\\
\dot{\rho}_6 &= \Gamma_7\rho_7 - (L_6 + A' + A)\rho_6 + a'\rho_2 + a\rho_4,
\end{aligned} \tag{49}$$

where the tunneling rates b, B, a, and A in one TLS are determined by Eqs. (29) and (23), and the primed tunneling rates for the second TLS are determined by similar expressions. The coefficients

$$\begin{aligned}
&L_0 = k^g_{10}, \quad L_2 = k^g_{32}, \quad L_4 = k^g_{54}, \quad L_6 = k^g_{76},\\
&L_1 = k^e_{10}, \quad L_3 = k^e_{32}, \quad L_5 = k^e_{54}, \quad L_7 = k^e_{76}
\end{aligned} \tag{50}$$

determine the rates of induced vertical transitions shown in Fig. 5.

Let us again turn to the case in which the laser pumping resonates with the frequency of the 1–0 transition. If light-induced tunneling is neglected, we can set $b = B = b' = B' = 0$. Then the set of equations (49) is written in the simplified form:

$$\begin{aligned}
\dot{\rho}_1 &= -\Gamma_1\rho_1 + L_0\rho_0,\\
\dot{\rho}_0 &= \Gamma_1\rho_1 - L_0\rho_0 - (a + a')\rho_0 + A\rho_2 + A'\rho_4,\\
\dot{\rho}_2 &= ap_0 - (A + a')\rho_2 + A'\rho_6,\\
\dot{\rho}_4 &= a'p_0 - (A' + a)\rho_4 + A\rho_6,\\
\dot{\rho}_6 &= a'p_2 + a\rho_4 - (A' + A)\rho_6.
\end{aligned} \tag{51}$$

These equations, like Eqs. (30), also have an electronic and tunneling block, with the tunneling block expressed in a more complicated form. Since the scheme suggested above is based on calculating separate solutions of these two blocks and combining these solutions, it can also be applied to find-

ing an approximate solution of the latter equations. The expressions for n_1 and n_0 are the same because the electronic block has remained unchanged. Only the slowly changing solution of the tunneling block is different. The approximate solution takes the form

$$\rho_1(t)=n_1(t)P_0(t), \quad \rho_2(t)=n_0(t)P_2(t),$$
$$\rho_4(t)=n_0(t)P_4(t), \quad \rho_6(t)=n_0(t)P_6(t),$$
$$\rho_0(t)=n_1(t)[1-P_0(t)]+n_0(t)P_0(t). \tag{52}$$

Here the expressions for the populations ρ_0 and ρ_1 of the two levels coupled by optical transitions are essentially unchanged. New expressions for the populations of levels 4 and 6, which are not directly involved in optical transitions, are calculated in a manner similar to that of the previously found population ρ_2. Obviously, the tunneling rates P_A satisfy the equations

$$\dot{P}_0=-(a+a')P_0+AP_2+A'P_4,$$
$$\dot{P}_2=-(A+a')P_2+aP_0+A'P_6,$$
$$\dot{P}_4=-(a+A')P_4+a'P_0+AP_6,$$
$$\dot{P}_6=-(A+A')P_6+a'P_2+aP_4, \tag{53}$$

whose solution is

$$P_0=p_0p_0', \quad P_2=p_2p_0', \quad P_4=p_0p_2', \quad P_6=p_2p_2', \tag{54}$$

where the coefficients are the rates characterizing each TLS and determined by the equations

$$p_0=-ap_0+Ap_2, \quad p_0'=-a'p_0'+A'p_2',$$
$$p_2=ap_0-Ap_2, \quad p_2'=a'p_0'-A'p_2'. \tag{55}$$

The rates satisfy the condition

$$P_0+P_2+P_4+P_6=(p_0+p_2)(p_0'+p_2')=1. \tag{56}$$

Equations (52) describe the desired solution of Eqs. (51) controlled by the optical pumping at the frequency of the 1–0 transition. We can similarly analyze the case in which the pumping frequency is resonant with the 3–2 transition. Then we have instead of Eqs. (52) the expressions

$$\rho_3(t)=n_3(t)P_2(t), \quad \rho_0(t)=n_2(t)P_0(t),$$
$$\rho_4(t)=n_2(t)P_4(t), \quad \rho_6(t)=n_2(t)P_6(t),$$
$$\rho_2(t)=n_3(t)[1-P_2(t)]+n_2(t)P_2(t), \tag{57}$$

where

$$n_3(t)=\frac{L_2}{\Gamma_3+L_2}\{1-\exp[-(\Gamma_3+L_2)t]\},$$
$$n_2(t)=1-n_3(t), \tag{58}$$

whose difference from Eqs. (52) is that subscripts 0 and 1 are replaced with 2 and 3.

Obviously, an increase in the number of TLSs only leads to a further increase in the size of the tunneling block without affecting the electronic one. Therefore, equations for the case in which a chromophore interacts with many TLSs can be easily obtained by trivially generalizing the latter equations. To this end, let us write the above formulas for the optical pumping at the frequency resonant with the $N \leftarrow M$ transition in the form

$$\rho_N(t)=n_N(t)P_M(t), \quad \rho_{M'}(t)=n_M(t)P_{M'}(t) \quad (M'\neq M),$$
$$\rho_M(t)=n_N(t)[1-P_M(t)]+n_M(t)P_M(t), \tag{59}$$

which also applies to the case in which a chromophore interacts with N_0 TLSs. In this case the probability $P_M(t)$, however, is a product of N_0 one-particle probabilities $p_j(t)$.

7. TWO-PHOTON CORRELATOR

The expression for the two-photon correlator contains the probability to find the excited electronic state, which is populated by absorption a laser photon. Consequently, when the laser resonates with the $N \leftarrow M$ transition, the expression for the two-photon correlator has the form

$$p(t)=\frac{\rho_N(t)}{T_1}. \tag{60}$$

By substituting Eq. (60) in the first equation in (59) and using the condition $L_M/\Gamma \ll 1$, we can express the two-photon correlator in the form

$$p(t)=\frac{n_N(t)P_M(t)}{T_1}\simeq\frac{L_MP_M}{T_1\Gamma_N}[1-\exp(-\Gamma_Nt)]$$
$$\simeq L_MP_M[1-\exp(-t/T_1)]. \tag{61}$$

The optical transitions shown in Fig. 5 and their corresponding Lorentzians $L_0, L_2, L_4, \ldots$ have different resonant frequencies. When the laser frequency is scanned, the main contribution results from the Lorentzian whose resonant frequency coincides with the laser frequency. With this fact in mind, we can transform Eq. (61) to

$$p(t)=k(\Delta_0,t,T)[1-\exp(-t/T_1)], \tag{62}$$

where

$$k(\Delta_0,t,T)=\sum_M L_MP_M(t) \tag{63}$$

is in fact the coefficient of light absorption by a molecule as a function of the frequency detuning Δ_0, temperature T, and time t measured with respect to the moment when the first photon of a pair was emitted. If the time tends to infinity, this formula describes the conventional coefficient of light absorption by a molecule interacting with photons and equilibrium TLSs. After substitution of L_M given by Eqs. (28) in the latter equation, it transforms to

$$k(\Delta_0,t,T)=\sum_{N,M} P_M(t)k_{NM}$$
$$=\chi^2\sum_{N,M} P_M(t)\langle N|M\rangle^2$$
$$\times\sum_{b,a}\rho_a\langle b|a\rangle^2\frac{1/T_1}{(\Delta_0+\Omega_{ba}+\Omega_{NM})^2+(1/2T_1)^2}$$

$$+\chi^2 \sum_{N,M} P_M(t)\langle N|M\rangle^2 \times \sum_{\beta,\alpha} \rho_\alpha \langle\beta|\alpha\rangle^2 \frac{1/T_1}{(\Delta_0+\Omega_{\beta\alpha}+\Omega_{NM})^2+(1/2T_1)^2}$$

$$=\chi^2 \sum_{B,A} w_A(t,T)\langle B|A\rangle^2 \times \frac{1/T_1}{(\Delta_0+\Omega_{BA})^2+(1+2T_1)^2}, \quad (64)$$

where the probability

$$w_A(t,T)=P_M(t)\rho_A(T)=P_M(t)\begin{Bmatrix}\rho_a\\ \rho_\alpha\end{Bmatrix} \quad (65)$$

of detecting A state of the tunnelon-phonon system is a function of both time and temperature.

The latter two equations take into account the effect of all TLSs, both in and out of equilibrium, on the optical transition through the overlap integrals $\langle N|M\rangle$ and probabilities $P_M(t)$. The time dependence of the probabilities is only determined by nonequilibrium TLSs through their tunneling rates $R=a+A$.

The shape of the optical absorption band was calculated using Eq. (64) with due account of equilibrium phonons and tunnelons[7,8] in the dynamic approach only based on the system Hamiltonian. Subsequent calculations[9–11] took into account the nonequilibrium condition of TLSs, which resulted in a time dependence of $P_M(t)$. It turned out that

$$k(\Delta_0,t,T)=2\chi^2 \int_{-\infty}^{\infty} I(\Delta_0,t,T,x)\exp\left(-\frac{|x|}{2T_1}\right)dx, \quad (66)$$

where the dipole correlator of the chromophore interacting with nonequilibrium TLSs is described by the following expression:[10,11]

$$I(\Delta_0,t,T,x)=\prod_j^{N_0} I_j(\Delta_0,t,T,x) =\exp\left[i(\Delta_0-\delta)x-\frac{|x|\gamma}{2}\right]\prod_j^{N_0}\{1-C_j(t,T) \times[1-\exp(-i\Delta_j x-|x|R_j)]\}, \quad (67)$$

where

$$C_j=\frac{\Delta_j}{\Delta_j-iR_j}p_j(t,T), \quad (68)$$

$$\delta=\sum_{j=1}^{N_0}\frac{R_j^2\Delta_j}{\Delta_j^2+R_j^2}p_j(1-p_j),$$

$$\frac{\gamma}{2}=\sum_{j=1}^{N_0}\frac{R_j^2\Delta_j^2}{\Delta_j^2+R_j^2}p_j(1-p_j), \quad (69)$$

$$p_j=\frac{1}{\exp(\hbar\varepsilon_j/kT)+1}[1-\exp(-R_j t)], \quad (70)$$

and $\hbar\varepsilon_j$, Δ_j, and R_j are the splitting between the levels, the change in the splitting due to excitation of the electronic system, and the relaxation rate constant of the jth TLS, respectively.

An important point is that the dipole correlator $I(x)$ given by Eq. (67) leads to the Lorentzian shape of a line, irrespective of the type of the electrostatic interaction between the chromophore and TLS. In point of fact, the lines of single molecules and spectral holes have Lorentzian shapes. On the other hand, the correlator $I(x)$ had been calculated many times on the basis of the stochastic approach,[12–16] and the line shapes obtained using $I(x)$ depended on the interaction type. This difference between the results can be ascribed to the fact that the tunnelon-phonon interaction was introduced in the earlier works[12–16] only partially through the finite lifetime $1/R$ of the tunnelon. The effect of this interaction on the amplitudes of electron-tunnelon transitions was not taken into consideration, which is equivalent to setting $C_j=p_j$ in Eq. (67). The exponentially decaying factor in front of the product sign in Eq. (67) was also not found in the stochastic theories.[12–16] These two differences lead to contradictions between the predictions of the two theories concerning the line shape and the effect of the type of interaction between the chromophore and TLS.

According to Eq. (67), based on the dynamic theory, the great number of TLSs located far from the chromophore and having small Δ_j do not contribute to the product over j because the function C_j cuts off the effective interaction at large distances. At the same time, the distant TLSs make a significant contribution to the dephasing constant $\gamma/2$. In the stochastic theories, on the contrary, the great number of remote TLSs make a decisive contribution to the product over j, which leads to the absorption band broadening. The broadening resulting from the product over j actually depends strongly on the type of interaction between the chromophore and TLS. The difference between the stochastic and dynamic approaches to the problem of line broadening is discussed in detail elsewhere.[11]

Formula (69) for $\gamma/2$ characterizes the rate of optical dephasing due to the interaction between the chromophore and all intrinsic TLSs. The full rate of the optical dephasing is given by the formula

$$\frac{1}{T_2(t,T)}=\frac{\gamma_{\rm ph}(T)}{2}+\frac{\gamma(t,T)}{2}+\frac{1}{2T_1}, \quad (71)$$

where the first term is due to the electron-phonon interaction and the second is due to the interaction with equilibrium and nonequilibrium intrinsic TLSs. Only the second term determines the time dependence of the optical dephasing, which is called spectral diffusion. It is a linear function of temperature and a logarithmic function of time.[11,15–17] The emergence of the spectral-diffusion term in the expression for the optical dephasing rate, hence in the expression for the two-photon correlator, is the main result of the calculation taking into account the molecule's interaction with many nonequilibrium TLSs.

Now let us apply general formula (62) to a chromophore whose optical absorption band consists of two resolved lines. This means that the chromophore strongly interacts with one

TLS in its vicinity, i.e., the product in Eq. (67) reduces to one term. The diagram of energy levels of this system is shown in Fig. 4. In this case

$$k(t,T)=L_0p_0+L_2p_2=2\chi^2[p_0(t)L(\omega)+p_2(t)L(\omega-\Delta)], \quad (72)$$

where

$$L(\omega)=\frac{1/T_2(t,T)}{\omega^2+1/T_2^2(t,T)}. \quad (73)$$

Here the frequency of the purely electronic transition is set to zero. The interaction with a great number of intrinsic TLSs shows up in the time dependence of T_2. The probabilities $p_j(t)$ are given by

$$p_2(t+t_0)=f+[p_2(t_0)-f]e^{-Rt},$$
$$p_0(t+t_0)=1-p_2(t+t_0). \quad (74)$$

Here $f=a/R=[\exp(\hbar\varepsilon/kT)+1]^{-1}$ is the population in thermal equilibrium.

Let the laser line frequency coincide with that of the 1–0 transition. Then the probability $p_2(t_0=0)=0$ at the initial time t_0, which is when the photon labeled by zero is detected. If the laser is tuned to the 1–0 line peak, when $\omega=0$ and the contribution of the second Lorentzian can be neglected, Eq. (62) for the two-photon correlator takes the form

$$p(t)=2\chi^2T_2(t,T)[1-\exp(-t/T_1)][1-f(1-e^{-Rt})]. \quad (75)$$

This correlator is a product of three functions of time. As was shown in earlier publications,[11,15–17] the component of the optical dephasing rate $\gamma(t,T)$ due to interaction with many intrinsic TLSs is described by a logarithmic function of time. Therefore we can set in Eq. (75)

$$\frac{1}{T_2}=\frac{100}{T_1}[1+10^{-2}\ln(R_2t)]. \quad (76)$$

A logarithmic broadening of this type of a spectral line generated by a molecule in a polymer was actually detected in experiments with persistent spectral holes.[16] In Eq. (76) $R_2\simeq10^{10}\,\mathrm{s}^{-1}$ is the largest tunneling rate in TLSs, and the factor in front of the logarithm determines the ratio between the line FWHM $\gamma(t,T)$ due to the spectral diffusion to the FWHM γ_{ph} due to interaction with phonons.

Figure 6 shows the two-photon correlator (75) plotted versus time t between emitted photons of one pair. The exponential relaxation of TLS in the chromophore's neighborhood shows up in Fig. 6 in the form of the smoothed step with a width of approximately one order on the logarithmic time scale, and the spectral diffusion results in the linear time dependence of the two-photon correlator on the logarithmic time scale.

Now let us discuss how the suggested theory interprets spectral line jumps illustrated by Fig. 1. Such a jump can be treated as burning out the spectral component corresponding to a single molecule. In fact, this process considered in terms of an ensemble of molecules can be treated as spectral hole burning in the inhomogeneously broadened spectral band of the ensemble of molecules. Persistent holes are burned out when tunneling transitions in the electronic ground state of the chromophore are suppressed, i.e., $a=A\simeq0$. We examine just such a case in excitation with the 1–0 laser transition. Then, given that $\rho_0(0)=1$ at the initial moment, when the first photon is detected, we derive from Eqs. (32) and (34) the following expression for the two-photon correlator:

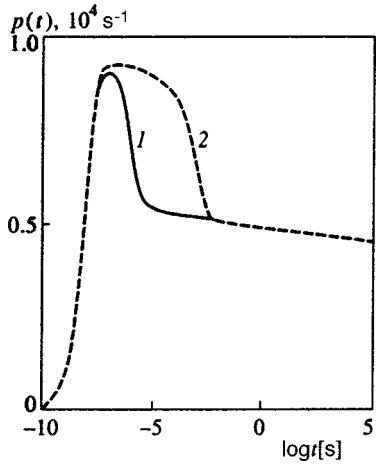

FIG. 6. Manifestation of the spectral diffusion and exponential relaxation due to jumps in TLS in the two-photon correlator: 1) $R=10^6\,\mathrm{s}^{-1}$; 2) $10^3\,\mathrm{s}^{-1}$.

$$p(t)=\frac{\rho_1(t)+\rho_3(t)}{T_1}\simeq L_0\exp(-\tilde{B}t). \quad (77)$$

When $\tilde{B}t_0\gg1$, the fluorescence intensity vanishes. Consequently, the spectral line jumps to a new position at a random time of order t_0. In fact, after the disappearance of the 1–0 line and tuning the laser to the frequency of the 3–2 transition, we can again detect emitted light, because in accordance with Eqs. (37) and (40), the two-photon correlator under such excitation has the form

$$p(t+t_0)=\frac{\rho_1(t+t_0)+\rho_3(t+t_0)}{T_1}\simeq L_2\exp(-\tilde{b}t). \quad (78)$$

In deriving the latter equation, we have taken into account that $\rho_2(t_0)=1$ at time t_0 of the onset of excitation at the 3–2 transition frequency. It follows from Eqs. (77) and (78) for the correlators that $1/\tilde{B}$ is the mean ''lifetime'' of the 1–0 transition line, and $1/\tilde{b}$ of the 3–2 transition line. In principle, these times can also be calculated by processing one-photon measurements discussed in Introduction. This procedure, however, is very complicated when analyzing situations more complex than that discussed here.

8. CONCLUSIONS

The main result of the reported investigation is development of the technique taking into account effects of a molecule's interaction with a great number of nonequilibrium intrinsic TLSs on the two-photon correlator. With this end in view, we have theoretically analyzed in detail in Secs. 3–6 the problem of how the interaction with nonequilibrium intrinsic TLSs can be introduced to the set of equations for the

density operator of a system including a two-level molecule, phonons, and long-lived excitations in TLSs. From the results of this analysis, we derived in Sec. 7 theoretical expression (62) for the two-photon correlator taking into account not only relaxation of TLS from the nearest vicinity of the molecule under consideration, but also relaxation of all intrinsic TLSs. All nontrivial information concerning relaxation of TLS is contained in function $k(\Delta_0,t,T)$, which tends to the conventional light absorption coefficient as $t \to \infty$.

The basic physical results of our investigation are illustrated by Fig. 6, which clearly shows that the two-photon correlator is controlled by both the exponential relaxation of TLS from the vicinity of a molecule and the relaxation of all intrinsic TLSs. The former process shows up in the form of broadened steps, and the latter in the logarithmic time dependence of optical dephasing rate $1/T_2$, i.e., the slow linear drop of the curve in Fig. 6.

By varying the frequency of the pumping laser, one can record, using the correlator technique, the profile of the extremely complicated optical band of a single molecule with a high accuracy, and by varying the time interval between two detected photons one can measure the rates of both the slow and fast TLS relaxation processes.

The technique based on measuring two-photon correlators in optical spectra of single molecules has obvious advantages over the conventional one-photon technique for measuring absorption spectra.

Using one-photon detection techniques, it is essentially impossible to measure the time dependence of the optical dephasing rate $1/T_2$, i.e., spectral diffusion, whereas this dependence can be easily determined using the expression for the two-photon correlator.

I thank L. B. Ershova for constructive criticism of this paper and the Russian Fund for Fundamental Research (Grant No. 97-02-17285) for financial support of the reported work.

*)e-mail: osad.ig@g23.relcom.ru

[1] *Single Molecule Optical Detection, Imaging and Spectroscopy*, by T. Basche, W. E. Moerner, M. Orrit, and U. Wild (eds.), VCH Verlagsgesellschaft mbH, Weinheim (1997).
[2] W. E. Moerner and T. Basche, Angew. Chem. **32**, 457 (1993).
[3] P. Thenio, A. B. Myers, and W. E. Moerner, J. Lumin. **56**, 1 (1993).
[4] L. Fleury, A. Zumbusch, M. Orrit, R. Brown, and J. Bernard, J. Lumin. **56**, 15 (1993).
[5] I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **113**, 1606 (1998) [JETP **86**, 875 (1998)].
[6] I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **98**, 1045 (1990) [Sov. Phys. JETP **71**, 583 (1990)].
[7] N. N. Zaitsev and I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **104**, 4042 (1993) [JETP **77**, 950 (1993).
[8] D. V. Donskoi, N. N. Zaitsev, and I. S. Osad'ko, Chem. Phys. **176**, 135 (1993).
[9] I. S. Osad'ko and N. N. Zaitsev, Chem. Phys. Lett. **257**, 82 (1996).
[10] I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **109**, 805 (1996) [JETP **82**, 434 (1996)].
[11] I. S. Osad'ko and N. N. Zaitsev, Chem. Phys. Lett. **293**, 138 (1998).
[12] R. Klauder and P. W. Anderson, Phys. Rev. **125**, 912 (1962).
[13] B. D. Laikhtman, Phys. Rev. B **31**, 490 (1985).
[14] M. A. Krivoglaz, Zh. Éksp. Teor. Fiz. **88**, 2171 (1985) [Sov. Phys. JETP **61**, 1284 (1985)].
[15] W. Breinl, J. Friedrich, and D. Haarer, J. Chem. Phys. **81**, 3915 (1984).
[16] K. Littau, Y. S. Bay, and M. D. Fayer, J. Chem. Phys. **92**, 4145 (1990).
[17] I. S. Osad'ko, Mol. Cryst.-Liq. Cryst. **291**, 45 (1996).

Translation provided by the Russian Editorial office.

The intermediate glassy phase in a ferroelectric with dislocations

P. N. Timonin*)

Physics Research Institute, Rostov State University, 334090 Rostov-on-Don, Russia
(Submitted 24 March 1999)
Zh. Éksp. Teor. Fiz. **116**, 986–1000 (September 1999)

This paper examines, in the mean-field approximation, the phase transitions in an elasto-isotropic cubic ferroelectric crystal with randomly distributed ring dislocations. It is found that a transition to the dipole-glass phase with chaotic, spontaneous unit-cell dipole moments, characterized by the Edwards–Anderson nonlocal parameter, precedes the transition to the ferroelectric phase. Also determined are the features of the thermodynamic parameters of the crystal at the phase-transition points. © *1999 American Institute of Physics.* [S1063-7761(99)01609-1]

1. INTRODUCTION

Numerous investigations of various phase transitions in real crystals (i.e., crystals with defects) have shown that the defects in the crystal structure have profound effects on the nature and features of the transitions. For instance, the presence of point defects in a crystal may change the critical exponents of a second-order transition if the specific-heat exponent of the ideal crystal is positive.[1,2] Extended defects also affect the critical behavior.[3–5] Furthermore, extended defects such as dislocations may give rise to local ordering in their vicinity above the temperature of a global second-order phase transition if there are short-range fluctuations of the order parameter.[6,7]

As noted by Dubrovskiĭ and Krivoglaz,[7] the appearance and growth near dislocations of ordered regions with a random sign of the order parameter appear as a smearing of the phase transition, a phenomenon often observed in experiments. Most often this smearing can be detected in ferroelectric crystals, where it manifests itself in the broadening of the peak in the dielectric constant. Generally speaking, however, the results of Nabutovskiĭ and Shapiro[6] and Dubrovskiĭ and Krivoglaz[7] cannot be applied to ferroelectric crystals, since there is a fairly strong dipole–dipole interaction in such crystals. At the same time, there is a high probability that in such crystals, too, inhomogeneous structures may be induced by dislocations above the point of transition to the ferroelectric phase. Qualitatively, the mechanism by which such structures form can be interpreted as the result of tuning of the directions of spontaneous local dipole moment to the intrinsic-electric-field fluctuations caused by random deformations of the crystal with the distributed dislocations. Here the nonparallel ordering of the local dipole moments is more advantageous as long as spontaneous polarization is small and the loss in the short-range energy of an inhomogeneity is not larger than the gain in the electrostatic energy, so that inhomogeneous structures may exists in a temperature interval near the temperature T_c of transition to the ferroelectric phase in the ideal crystal. More than that, the directions of polarization in such structures may prove to be correlated over the entire bulk of the crystal due to the long-range nature of polarization fluctuations. This occurrence of an inhomogeneous distribution of polarization near T_c is an indication of a transition to the intermediate dipole-glass phase accompanied by specific anomalies of thermodynamic quantities.[8] In this paper we will discuss, in the mean-field approximation, the possibility of such an intermediate glassy phase developing in the simplest case of an elasto-isotropic cubic ferroelectric crystal with a low concentration of randomly distributed ring dislocations. We will find that in this case the transition to the ferroelectric phase is indeed preceded by a transition to the dipole-glass phase with inhomogeneous spontaneous polarization characterized by the Edwards–Anderson nonlocal parameter

2. THE REPLICA FORMALISM FOR A FERROELECTRIC WITH DISLOCATIONS

The Ginzburg–Landau thermodynamic potential for a cubic ferroelectric with dislocations has the form

$$F=\int_V d\mathbf{r}\left\{\frac{1}{2}[\tau\mathbf{P}^2(\mathbf{r})-l^2\mathbf{P}(\mathbf{r})\cdot\Delta\mathbf{P}(\mathbf{r})]+\frac{u}{4}(\mathbf{P}^2(\mathbf{r}))^2+\frac{v}{4}\sum_{i=1}^{3}P_i^4\right\}+\frac{1}{2}\int_V d\mathbf{r}\int_V d\mathbf{r}'\mathbf{P}(\mathbf{r})\hat{D}\times(\mathbf{r}-\mathbf{r}'+\mathbf{u}(\mathbf{r})-\mathbf{u}(\mathbf{r}'))\mathbf{P}(\mathbf{r}'). \quad (1)$$

where $\mathbf{P}(\mathbf{r})$ is the polarization, $\mathbf{u}(\mathbf{r})$ represent the atomic displacement caused by ring dislocations randomly distributed in the crystal, and $\tau=(T-T_c)/T_0$. Below we assume that the coupling constants u and v are small, with

$$u\ll v. \quad (2)$$

This condition means that in an ideal crystal there is a transition to the rhombohedral ferroelectric phase with an isotropic dielectric susceptibility

$$\chi_{ij}=\frac{\partial P_i}{\partial E_j}=\chi\delta_{ij}.$$

The fact simplifies calculations in the ferroelectric phase but is not needed in determining whether there exists an inter-

mediate glassy phase. In what follows, $\hat{D}(\mathbf{r})$ is the dipole–dipole interaction tensor, which in the absence of conductors has the form[9,10]

$$\hat{D}(\mathbf{r}) = -\nabla\nabla\frac{1}{r}.$$

The tensor $\hat{D}(\mathbf{r})$ has the same form when there are conductors on the surface of the ferroelectric but at distances much smaller than the crystal size, $r \ll L$. In this case we have[10]

$$\int_V d\mathbf{r}\hat{D}(\mathbf{r}) = 0. \tag{3}$$

We take this case, where the transition in an ideal crystal is to the monodomain state, to be the simplest one. But if there are no conductors, the transition is to the polydomain state, whose description is somewhat more complicated.[10]

Here we also consider only the effect of dislocations on the long-range interaction, assuming that this effect on the short-range interaction is much weaker in view of the usual smallness of the electrostrictional constants.

Below we will be interested in the behavior of the polarization correlators over large distances, much larger than the lattice constant a and the average size of the dislocation loops, since it determines the behavior of the thermodynamic quantities near phase transitions. At such distances from the dislocation loops, the displacements $\mathbf{u}(\mathbf{r})$ can be written[11]

$$u_i(\mathbf{r}) = -\sum_m \partial_k C_{ij}(\mathbf{r}-\mathbf{r}_m)\lambda_{jklm}d^a_{lm}. \tag{4}$$

Here the vector $\mathbf{r}_m$ determines the position of an individual dislocation, λ_{jklm} is tensor of the elastic moduli of the crystal, $C_{ij}(\mathbf{r})$ is the Green's function of the equations of the theory of elasticity, $\lambda_{iklm}\partial_k\partial_l C_{mj}(\mathbf{r}) = \delta(\mathbf{r})\delta_{ij}$, and $d^a_{lm} = S^a_l b^a_m$ is the dislocation moment of a closed dislocation ($\mathbf{S} = \int_{S_D} d\mathbf{S}$, where S_D is the surface stretched over the dislocation contour and $\mathbf{b}$ is the Burgers vector of the dislocation).

We seek the pair correlators of the polarization Fourier transforms

$$\mathbf{P}(\mathbf{k}) = \frac{1}{\sqrt{V}}\int_V d\mathbf{r}\mathbf{P}(\mathbf{r})\exp\{i\mathbf{k}\cdot\mathbf{r}\},$$

averaged over the positions of the dislocations and the their momenta; more precisely, we are looking for the inhomogeneous susceptibility

$$G_{ij}(\mathbf{k}) = \beta\langle P_i(\mathbf{k})P_j(-\mathbf{k})\rangle_{T,C}, \quad \beta = T^{-1}, \tag{5}$$

and the Edwards–Anderson nonlocal parameter

$$Q_{ij}(\mathbf{k}) = \langle\langle P_i(\mathbf{k})\rangle_T\langle P_j(-\mathbf{k})\rangle_T\rangle_C. \tag{6}$$

Here the subscript T stands for Gibbs averaging with the potential F [see Eq. (1)], and the subscript C indicates averaging over the random distribution of the dislocations.

We must also find the average value of the density of the equilibrium thermodynamic potential, $\Phi = -(T/V)\ln\langle Z\rangle_C$, with $Z = \int D\mathbf{P}\exp\{-\beta F\}$. We will use the replica method,[8] according to which

$$\Phi = -\frac{T}{V}\lim_{n\to 0}\frac{1}{n}\ln\langle Z^n\rangle_C, \tag{7}$$

i.e., we must average Z^n for integral values of n, continue the resulting expression to nonintegral values of n, and then let n go to zero. For integral values of n, with allowance for (2) we get

$$\langle Z^n\rangle = \int D\mathbf{P}_\alpha(\mathbf{r})\exp\{-\beta F_n\}, \tag{8}$$

$$F_n = \sum_{\alpha=1}^{n}\int d\mathbf{r}\left\{\frac{1}{2}[\tau\mathbf{P}^2_\alpha(\mathbf{r}) - l^2\mathbf{P}_\alpha(\mathbf{r})\cdot\Delta\mathbf{P}_\alpha(\mathbf{r})] + \frac{\nu}{4}\sum_{i=1}^{3}P^4_{i\alpha}\right\} + F_{\rm dis}, \tag{9}$$

$$F_{\rm dis} = -T\ln\left\langle \exp\left(\sum_{\alpha=1}^{n}\frac{1}{2T}\int d\mathbf{r}\int d\mathbf{r}'\mathbf{P}_\alpha(\mathbf{r})\hat{D}\times(\mathbf{r}-\mathbf{r}'+\mathbf{u}(\mathbf{r})-\mathbf{u}(\mathbf{r}'))\mathbf{P}_\alpha(\mathbf{r}')\right)\right\rangle_C. \tag{10}$$

We are interested in the case of large differences $\mathbf{r}-\mathbf{r}'$ and small dislocation number densities $n_{\rm dis} = N_{\rm dis}/V$. Then the quantities $\mathbf{u}(\mathbf{r})$, which decrease at large distances, will be small, so that if we expand $F_{\rm dis}$ in the small differences $\mathbf{u}(\mathbf{r})-\mathbf{u}(\mathbf{r}')$, we get

$$F_{\rm dis} = -\frac{1}{2VT}\sum_m\int d\mathbf{r}_m\left\langle\left(\sum_{\alpha=1}^{n}\int d\mathbf{r}\int d\mathbf{r}'\mathbf{P}_\alpha(\mathbf{r})\times[\mathbf{u}_m(\mathbf{r}-\mathbf{r}_m)\nabla]\hat{D}(\mathbf{r}-\mathbf{r}')\mathbf{P}_\alpha(\mathbf{r}')\right)^2\right\rangle_C. \tag{11}$$

where $\mathbf{u}_m(\mathbf{r}-\mathbf{r}_m)$ is the contribution of a single dislocation to the atomic displacements [see Eq. (4)], and averaging over the dislocation positions, in view of the assumed smallness of $n_{\rm dis}$, reduces to integrating over the vectors $\mathbf{r}_m$ that specify these positions.

If in (11) we go over to Fourier transforms and integrate over $\mathbf{r}_m$, we arrive at the final expression:

$$F_{\rm dis} = -\frac{n_{\rm dis}}{2T}V^2\sum_{\alpha,\beta=1}^{n}\int\frac{d\mathbf{k}\,d\mathbf{q}_1\,d\mathbf{q}_2}{(2\pi)^9}\times[\mathbf{P}_\alpha(\mathbf{k}-\mathbf{q}_1)\hat{D}(\mathbf{q}_1)\mathbf{P}_\alpha(\mathbf{q}_1)][\mathbf{P}^*_\beta(\mathbf{k}-\mathbf{q}_2)\times\hat{D}(\mathbf{q}_2)\mathbf{P}^*_\beta(\mathbf{q}_2)][\mathbf{q}_1\hat{\Delta}(\mathbf{k})\mathbf{q}_2], \tag{12}$$

$$\Delta_{ij}(\mathbf{k}) = C_{ij}(\mathbf{k})k_l\langle\tilde{d}_{kl}\tilde{d}_{mn}\rangle_C k_m C_{mj}(\mathbf{k}), \quad \tilde{d}_{ij} = \lambda_{ijkl}d_{kl}, \tag{13}$$

$$D_{ij}(\mathbf{k}) = \frac{4\pi k_i k_j}{k^2}, \quad kL \gg 1. \tag{14}$$

Note that integration over the wave vectors in (12) is done in the first Brillouin zone, $|k_i| < \pi/a$.

It is natural to consider such distributions of dislocation moments that on the average preserve the cubic symmetry of the crystal, so that the mean-square dislocation-moment ten-

sor $\langle d_{ij}d_{kl}\rangle_C$ and the tensor $\langle \tilde{d}_{kl}\tilde{d}_{mn}\rangle_C$, which enters into (13), have the same finite components as λ_{ijkl}. Assuming that the crystal is elasto-isotropic, so that

$$\lambda_{ijkl}=\lambda_{12}\delta_{ij}\delta_{kl}+\frac{1}{2}(\lambda_{11}-\lambda_{12})(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}),$$

$$\langle d_{ij}d_{kl}\rangle=\langle d_{11}d_{22}\rangle\delta_{ij}\delta_{kl}+\frac{1}{2}\langle d_{11}^2-d_{11}d_{22}\rangle(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}),$$

we can represent $\Delta_{ij}(\mathbf{k})$ in the form

$$\Delta_{ij}(\mathbf{k})=\Delta_l k^{-4}k_ik_j+\Delta_t k^{-4}(k^2\delta_{ij}-k_ik_j),$$

$$\Delta_l=\langle d_{11}^2\rangle\left(1+2\frac{\lambda_{12}^2}{\lambda_{11}^2}\right)+2\langle d_{11}d_{12}\rangle(2\lambda_{11}+\lambda_{12})\frac{\lambda_{12}}{\lambda_{11}^2},$$

$$\Delta_t=2\langle d_{11}^2-d_{11}d_{22}\rangle.$$

To be able to describe the transition into the ferroelectric phase, we introduce the thermodynamic potential $\Phi_n(\mathbf{P}_\alpha)$, which can be obtained from F_n by isolating the homogeneous polarization $\mathbf{P}_\alpha\equiv V^{-1}\int_V d\mathbf{r}\mathbf{P}_\alpha(\mathbf{r})$ in the expression for F_n:

$$\Phi_n(\mathbf{P}_\alpha)=-\frac{T}{V}\ln\prod_{\alpha=1}^{n}d\mathbf{P}_\alpha\delta\left(\mathbf{P}_\alpha-\frac{1}{V}\int_V d\mathbf{r}\mathbf{P}_\alpha(\mathbf{r})\right)\times\exp\{-\beta F_n\}, \tag{15}$$

so that

$$\Phi=-\lim_{n\to0}\frac{T}{nV}\ln\int\prod_{\alpha=1}^{n}d\mathbf{P}_\alpha\exp\{-\beta V\Phi_n(\mathbf{P}_\alpha)\}\approx\lim_{n\to0}\frac{1}{n}\min_{P_\alpha}\mathbf{P}_\alpha. \tag{16}$$

Note that Eq. (15) implies that the equilibrium value $\mathbf{P}_\alpha^0$ where $\Phi_n(\mathbf{P}_\alpha)$ attains its minimum determines the average value of homogeneous spontaneous polarization:

$$\langle\mathbf{P}\rangle_{T,C}=\lim_{n\to0}\frac{1}{n}\sum_{\alpha=1}^{n}\mathbf{P}_\alpha^0. \tag{17}$$

We can find $\mathbf{P}_\alpha^0$ by solving the equation of state

$$\frac{\partial\Phi(\mathbf{P}_\alpha)}{\partial\mathbf{P}_\alpha}=0. \tag{18}$$

Assuming that in (9) $\mathbf{P}_\alpha(\mathbf{r})=\mathbf{P}_\alpha+\mathbf{P}'_\alpha(\mathbf{r})$, we obtain

$$\Phi_n(\mathbf{P}_\alpha)=\sum_{\alpha=1}^{n}\left(\frac{\tau}{2}\mathbf{P}_\alpha^2+\frac{\nu}{4}\sum_{i=1}^{3}P_{i\alpha}^4\right)-\frac{T}{V}\ln\int D\mathbf{P}'_\alpha(\mathbf{r})\exp\{-\beta F_n(\mathbf{P}_\alpha,\mathbf{P}'_\alpha(\mathbf{r}))_n\}, \tag{19}$$

$$F_n(\mathbf{P}_\alpha,\mathbf{P}'_\alpha(\mathbf{r}))=F_n(\mathbf{P}'_\alpha(\mathbf{r}))+\delta F_n(\mathbf{P}_\alpha,\mathbf{P}'_\alpha(\mathbf{r})). \tag{20}$$

Here $F_n[\mathbf{P}'_\alpha(\mathbf{r})]$ has the same form as F_n in (9), and δF_n in the presence of conductors [see (3)] is given by the following formula (which is valid in the lowest order in $\mathbf{P}'_\alpha(\mathbf{k})$ sufficient in the mean-field region):

$$\delta F_n(\mathbf{P}_\alpha,\mathbf{P}'_\alpha(\mathbf{r}))=\frac{3}{2}V\int\frac{d\mathbf{k}}{(2\pi)^3}\left(\nu\sum_{i,\alpha}P_{i\alpha}^2\left|P'^2_{i\alpha}(\mathbf{k})\right|-\beta wk^{-4}\left|\sum_{\alpha=1}^{n}(\mathbf{k}\cdot\mathbf{P}_\alpha)(\mathbf{k}\cdot\mathbf{P}'_\alpha(\mathbf{k}))\right|^2\right), \tag{21}$$

where $w\equiv(4\pi)^2n_{\rm dis}\Delta_l/3$. In deriving (21) we ignored the contribution of the region where $kL\leqslant1$ to the integral over $\mathbf{k}$ and used (14).

If we employ (15) for the potential $\Phi_n(\mathbf{P}_\alpha)$, we can find not only the equilibrium thermodynamic potential Φ [Eq. (16)] and the average spontaneous polarization (17) but also other thermodynamic quantities. For instance, the definition of a correlator,

$$G_{ij}^{\alpha\beta}=\beta V\langle P_{i\alpha}P_{j\beta}\rangle_{\Phi_n}=\left[\frac{\partial^2\Phi_n(\mathbf{P}_\alpha)}{\partial P_{i\alpha}\partial P_{j\beta}}\right]^{-1}, \tag{22}$$

makes it possible to find the homogeneous dielectric susceptibility

$$\chi_{ij}=\frac{\partial P_i}{\partial E_j}=\lim_{n\to0}\frac{1}{n}\sum_{\alpha=1}^{n}G_{ij}^{\alpha\alpha} \tag{23}$$

and the homogeneous glassy correlator

$$S_{ij}=V\langle\langle P_i\rangle_T\langle P_j\rangle_T\rangle_C=\int_V d\mathbf{r}\langle\langle P_i(\mathbf{r})\rangle_T\langle P_j(0)\rangle_T\rangle_C=\lim_{n\to0}\frac{T}{n(n-1)}\sum_{\alpha\neq\beta}^{n}G_{ij}^{\alpha\beta}. \tag{24}$$

More than that, we can calculate the contribution of polarization fluctuations to the crystal specific heat:

$$\delta C=-T\frac{\partial^2\Phi}{\partial T^2}.$$

By differentiating the first equality in (16) with respect to T we get

$$\delta C=\lim_{n\to0}\frac{T}{nT_0^2}\left(\frac{\partial^2\Phi_n(\mathbf{P}_\alpha)}{\partial\tau\partial P_{i\alpha}}\times G_{ij}^{\alpha\beta}\frac{\partial^2\Phi_n(\mathbf{P}_\alpha)}{\partial\tau\partial P_{j\beta}}-\frac{\partial^2\Phi_n(\mathbf{P}_\alpha)}{\partial\tau^2}\right). \tag{25}$$

The potential $\Phi_n(\mathbf{P}_\alpha)$ can be found in the lowest order in the interaction by calculating, for $kL\gg1$, the replica correlator

$$G_{ij}^{\alpha\beta}(\mathbf{k})=\beta\langle P_i'^{\,\alpha}(\mathbf{k})P_j'^{\,\beta}(\mathbf{k})\rangle, \tag{26}$$

where we have assume that there is Gibbs averaging with the potential $F_n[\mathbf{P}_\alpha,\mathbf{P}'_\alpha(\mathbf{r})]$ [Eq. (20)] at the equilibrium value $\mathbf{P}_\alpha^0$ at which $\Phi_n(\mathbf{P}_\alpha)$ is at its minimum. Indeed, differentiating (7) with respect to τ, we get

$$\frac{\partial\Phi(\mathbf{P}_\alpha)}{\partial\tau}=\frac{1}{2}\sum_{\alpha=1}^{n}\mathbf{P}_\alpha^2+\frac{T}{2}\int\frac{d\mathbf{k}}{(2\pi)^3}\mathrm{Tr}\,\hat{G}(\mathbf{k}),\tag{27}$$

where $\mathrm{Tr}\hat{G}(\mathbf{k})$ is the trace of the replica correlator (26) regarded as a matrix of dimension $3n$. According to (20) and (21), the correlator $G_{ij}^{\alpha\beta}(\mathbf{k})$ can be written

$$[\hat{G}(\mathbf{k})^{-1}]_{ij}^{\alpha\beta}=[(\tau+l^2k^2+3\nu P_{i\alpha}^2)\delta_{ij}+D_{ij}(\mathbf{k})]\delta_{\alpha\beta}-3\beta w k^{-4}k_ik_j(\mathbf{k}\cdot\mathbf{P}_\alpha)(\mathbf{k}\cdot\mathbf{P}_\beta)-\Sigma_{ij}^{\alpha\beta}(\mathbf{k}),\tag{28}$$

where the dependence of the self-energy part $\Sigma_{ij}^{\alpha\beta}(\mathbf{k})$ on τ can be ignored, so that integration in (27) yields (to lowest order in the interaction)

$$\Phi_n(\mathbf{P}_\alpha)=\sum_{\alpha=1}^{n}\left(\frac{\tau}{2}\mathbf{P}_\alpha^2+\frac{\nu}{4}\sum_{i=1}^{3}P_{i\alpha}^4\right)-\frac{T}{2}\int\frac{d\mathbf{k}}{(2\pi)^4}\mathrm{Tr}\ln\hat{G}(\mathbf{k}).\tag{29}$$

Also in the lowest order in the interaction, Eqs. (9), (12), and (20) yield

$$\begin{aligned}\Sigma_{ij}^{\alpha\beta}(\mathbf{k})=&-3\nu T\delta_{ij}\delta_{\alpha\beta}\int\frac{d\mathbf{q}}{(2\pi)^3}G_{ii}^{\alpha\beta}(\mathbf{q})+n_{\mathrm{dis}}\\&\times\int\frac{d\mathbf{q}}{(2\pi)^3}\{[\mathbf{k}\hat{\Delta}(\mathbf{k}+\mathbf{q})\mathbf{k}]D_{ik}(\mathbf{k})G_{kl}^{\alpha\beta}(\mathbf{q})\\&\times D_{lj}(\mathbf{k})+[\mathbf{k}\hat{\Delta}(\mathbf{k}+\mathbf{q})\mathbf{q}]\\&\times[D_{ik}(\mathbf{k})G_{kl}^{\alpha\beta}(\mathbf{q})D_{lj}(\mathbf{q})\\&+D_{ik}(\mathbf{q})G_{kl}^{\alpha\beta}(\mathbf{q})D_{lj}(\mathbf{k})]+[\mathbf{q}\hat{\Delta}(\mathbf{k}\\&+\mathbf{q})\mathbf{q}]D_{ik}(\mathbf{q})G_{kl}^{\alpha\beta}(\mathbf{q})D_{lj}(\mathbf{q})\}.\end{aligned}\tag{30}$$

Formulas (28) and (30) comprise an equation for $G_{ij}^{\alpha\beta}(\mathbf{k})$ in the mean-field region in the case of weak interaction.

According to (18), (28), and (29), the equation of state for $\mathbf{P}_\alpha^0$ has the form

$$\frac{\partial\Phi(\mathbf{P}_\alpha)}{\partial P_{i\alpha}}=\tau P_{i\alpha}+\nu P_{i\alpha}^3+3T\int\frac{d\mathbf{k}}{(2\pi)^3}[\nu G_{ii}^{\alpha\alpha}(\mathbf{k})P_{i\alpha}-\beta w k^{-4}k_ik_lG_{lm}^{\alpha\beta}(\mathbf{k})k_m(\mathbf{k}\cdot\mathbf{P}_\beta)]=0,\tag{31}$$

while for the derivatives in (22) and (25) we have

$$\frac{\partial^2\Phi_n(\mathbf{P}_\alpha)}{\partial\tau^2}=-\frac{T}{2}\int\frac{d\mathbf{k}}{(2\pi)^3}\mathrm{Tr}\,\hat{G}^2(\mathbf{k}),\tag{32}$$

$$\frac{\partial^2\Phi(\mathbf{P}_\alpha)}{\partial P_{i\alpha}\partial P_{j\beta}}=(\tau+3\nu P_{i\alpha}^2)\delta_{ij}\delta_{\alpha\beta}+3T\times\int\frac{d\mathbf{k}}{(2\pi)^3}[\nu G_{ii}^{\alpha\alpha}(\mathbf{k})\delta_{ij}\delta_{\alpha\beta}-\beta w k^{-4}k_ik_jk_lG_{lm}^{\alpha\beta}(\mathbf{k})k_m],\tag{33}$$

$$\frac{\partial^2\Phi(\mathbf{P}_\alpha)}{\partial P_{i\alpha}\partial\tau}\approx P_{i\alpha}.\tag{34}$$

Thus, the solution of Eqs. (28) and (30) for $G_{ij}^{\alpha\beta}(\mathbf{k})$ provides a complete description of the thermodynamics of the crystal in the mean-field region. We can also find the inhomogeneous susceptibility $G_{ij}(\mathbf{k})$ [Eq. (5)] and the Edwards–Anderson nonlocal parameter $Q_{ij}(\mathbf{k})$ [Eq. (6)], since[8]

$$G_{ij}(\mathbf{k})=\lim_{n\to0}\frac{1}{n}\sum_{\alpha=1}^{n}G_{ij}^{\alpha\alpha}(\mathbf{k}),\tag{35}$$

$$Q_{ij}(\mathbf{k})=\lim_{n\to0}\frac{1}{n(n-1)}\sum_{\alpha\neq\beta}^{n}G_{ij}^{\alpha\beta}(\mathbf{k}).\tag{36}$$

3. CORRELATORS AND THE EQUATIONS OF STATE IN THE MEAN-FIELD APPROXIMATION

As noted earlier, when condition (2) is met, in an ideal crystal there is a transition to the rhombohedral ferroelectric phase. We will seek the replica-symmetric solutions of the equation of state (31). Accordingly, we will assume $P_{\alpha i}=P$, where, in agreement with (17), $P=\langle P_i\rangle_{T,C}$. Then the solution of Eqs. (28) and (30) for $kL\gg1$ can be written as the sum of the longitudinal and transverse parts:

$$G_{ij}^{\alpha\beta}(\mathbf{k})=G_l^{\alpha\beta}(\mathbf{k})k^{-2}k_ik_j+G_t^{\alpha\beta}(\mathbf{k})k^{-2}(k^2\delta_{ij}-k_ik_j),$$

where $G_s^{\alpha\beta}(\mathbf{k})$, $s=l,t$, has the form

$$[\hat{G}_s^{-1}(\mathbf{k})]^{\alpha\beta}=(\tau+3\nu P^2+l^2k^2+4\pi\delta_{sl})\delta_{\alpha\beta}-3\beta w k^{-2}(\mathbf{k}\cdot\mathbf{P})^2\delta_{sl}-\Sigma_s^{\alpha\beta}(\mathbf{k})\tag{37}$$

with the following self-energy parts:

$$\begin{aligned}\Sigma_l^{\alpha\beta}(\mathbf{k})=&-\nu T\delta_{\alpha\beta}\int\frac{d\mathbf{q}}{(2\pi)^3}[G_l^{\alpha\alpha}(\mathbf{q})+2G_t^{\alpha\alpha}(\mathbf{q})]\\&+(4\pi)^2n_{\mathrm{dis}}\int\frac{d\mathbf{q}}{(2\pi)^3}[\Delta_lx^2G_l^{\alpha\beta}(\mathbf{q})\\&+R(\mathbf{k},\mathbf{q})G_t^{\alpha\beta}(\mathbf{q})],\end{aligned}\tag{38}$$

$$\begin{aligned}\Sigma_t^{\alpha\beta}(\mathbf{k})=&-\nu T\delta_{\alpha\beta}\int\frac{d\mathbf{q}}{(2\pi)^3}[G_l^{\alpha\alpha}(\mathbf{q})+2G_t^{\alpha\alpha}(\mathbf{q})]\\&+(4\pi)^2n_{\mathrm{dis}}\int\frac{d\mathbf{q}}{(2\pi)^3}R(\mathbf{q},\mathbf{k})G_l^{\alpha\beta}(\mathbf{q}),\end{aligned}\tag{39}$$

$$R(\mathbf{k},\mathbf{q})=(1-x^2)\frac{\Delta_t[k^2q^2-(\mathbf{k}\cdot\mathbf{q})^2]+\Delta_l(k^2+\mathbf{k}\cdot\mathbf{q})^2}{|\mathbf{k}+\mathbf{q}|^4},\tag{40}$$

where $x\equiv\mathbf{k}\cdot\mathbf{q}/kq$. Equations (37)–(40) have replica-symmetric solutions of the form

$$G_s^{\alpha\beta}(\mathbf{k})=G_s^{11}(\mathbf{k})\delta_{\alpha\beta}+G_s^{12}(\mathbf{k})(1-\delta_{\alpha\beta}).\tag{41}$$

According to (35) and (36), with such solutions the correlators $G_{ij}(\mathbf{k})$ [Eq. (5)] and $Q_{ij}(\mathbf{k})$ [Eq. (6)], which are of interest to us, have the form

$$G_{ij}(\mathbf{k})=G_l(\mathbf{k})k^{-2}k_ik_j+G_t(\mathbf{k})k^{-2}(k^2\delta_{ij}-k_ik_j),$$

$$Q_{ij}(\mathbf{k})=Q_l(\mathbf{k})k^{-2}k_ik_j+Q_t(\mathbf{k})k^{-2}(k^2\delta_{ij}-k_ik_j),$$

$$G_s(\mathbf{k})=\lim_{n\to0}G_s^{11}(\mathbf{k}),\quad Q_s(\mathbf{k})=T\lim_{n\to0}G_s^{12}(\mathbf{k}).$$

Substituting (41) in (37)–(39) and letting n go to zero, we arrive at equations for $G_s(\mathbf{k})$ and $Q_s(\mathbf{k})$, $s=l,t$, for $kL \gg 1$:

$$Q_s(\mathbf{k}) = G_s'^2(\mathbf{k})[3wk^{-2}(\mathbf{k}\cdot\mathbf{P})^2\delta_{sl} + \Sigma_s^{12}(\mathbf{k})], \tag{42}$$

$$G_s(\mathbf{k}) = G_s'(\mathbf{k}) + \beta Q_s(\mathbf{k}), \tag{43}$$

$$G_s'^{-1}(\mathbf{k}) = \tau + 3vP^2 + l^2k^2 + 4\pi\delta_{sl} - \Sigma_s^{11}(\mathbf{k}) + \Sigma_s^{12}(\mathbf{k}), \tag{44}$$

$$\Sigma_l^{11}(\mathbf{k}) = -vT\int \frac{d\mathbf{q}}{(2\pi)^3}[G_l(\mathbf{q}) + 2G_t(\mathbf{q})] + (4\pi)^2 n_{\rm dis}\int \frac{d\mathbf{q}}{(2\pi)^3}[\Delta_l x^2 G_l(\mathbf{q}) + R(\mathbf{k},\mathbf{q})G_t(\mathbf{q})], \tag{45}$$

$$\Sigma_l^{12}(\mathbf{k}) = (4\pi)^2 n_{\rm dis}\int \frac{d\mathbf{q}}{(2\pi)^3}[\Delta_l x^2 Q_l(\mathbf{q}) + R(\mathbf{k},\mathbf{q})Q_t(\mathbf{q})], \tag{46}$$

$$\Sigma_t^{11}(\mathbf{k}) = -vT\int \frac{d\mathbf{q}}{(2\pi)^3}[G_l(\mathbf{q}) + 2G_t(\mathbf{q})] + \frac{(4\pi)^2}{2} n_{\rm dis}\int \frac{d\mathbf{q}}{(2\pi)^3} R(\mathbf{q},\mathbf{k})G_l(\mathbf{q}), \tag{47}$$

$$\Sigma_t^{12}(\mathbf{k}) = (4\pi)^2 n_{\rm dis}\int \frac{d\mathbf{q}}{(2\pi)^3} R(\mathbf{q},\mathbf{k})Q_l(\mathbf{q}). \tag{48}$$

We seek the spherically symmetric solutions $G_s(\mathbf{k}) = G_s(k)$ and $Q_s(\mathbf{k}) = Q_s(k)$ near the transition, $\tau \to 0$, for small $k \leqslant \sqrt{\tau/l}$, so that the expressions (45)–(48) for the self-energy parts can be simplified substantially by employing the fact that $G_t'(q)$ rapidly increases as $q \to 0$. Hence in integrals where $G_t'^2(q)$ is present the domain of integration is limited to small $q \leqslant \sqrt{\tau/l}$, so that, in particular, we have

$$\begin{aligned}\Sigma_l^{12}(k) &= wB_l + (4\pi)^2 n_{\rm dis}\int \frac{d\mathbf{q}}{(2\pi)^3} R(\mathbf{k},\mathbf{q})Q_t(\mathbf{q}) \\ &= wB_l + \frac{(4\pi)^4 n_{\rm dis}^2}{2}\int \frac{d\mathbf{q}\,d\mathbf{p}}{(2\pi)^6} R(\mathbf{k},\mathbf{q})G_t'^2(q) \\ &\quad \times R(\mathbf{p},\mathbf{q})Q_t(p) \approx wB_l + \frac{(4\pi)^4 n_{\rm dis}^2}{2} \\ &\quad \times \int \frac{d\mathbf{q}}{(2\pi)^3} R(\mathbf{k},\mathbf{q})G_t'^2(q) \\ &\quad \times \int \frac{d\mathbf{p}}{(2\pi)^3} R(\mathbf{p},0)Q_t(p) \\ &= wB_l\left[1 + (4\pi)^2 n_{\rm dis}\int \frac{d\mathbf{q}}{(2\pi)^3} R(\mathbf{k},\mathbf{q})G_t'^2(q)\right],\end{aligned} \tag{49}$$

where B_s with s equal to l or t is the Edwards–Anderson ''integral'' parameter Q_s defined according to (36) and (42):

$$B_s \equiv \int \frac{d\mathbf{q}}{(2\pi)^3} Q_s(q).$$

More than that, for small k we have

$$\Sigma_l^{11}(k) \approx \Sigma_t^{11}(k) \approx -vT(A_l + 2A_t) + wA_l, \tag{50}$$

$$\Sigma_t^{12}(k) \approx wB_l, \tag{51}$$

where

$$A_s \equiv \int \frac{d\mathbf{q}}{(2\pi)^3} G_s(q) = \Pi_{1s} + \beta B_s,$$

$$\Pi_{ms} \equiv \int \frac{d\mathbf{q}}{(2\pi)^3} G_s'^m(q).$$

Thus,

$$G_l'^{-1}(k) = \theta + l^2k^2 + 4\pi, \quad G_t'^{-1}(k) = \theta + l^2k^2, \tag{52}$$

$$\theta = \tau' + v(B_l + 2B_t + 3P^2),$$

$$\tau' \equiv \tau + vT(\Pi_{1l} + 2\Pi_{1t}) - w\Pi_{1l}. \tag{53}$$

As $\theta \to 0$,

$$\Pi_{1l} \approx \frac{1}{2\pi l^2 a}\left(1 - \frac{2a}{\sqrt{\pi}l}\arctan\frac{\sqrt{\pi}l}{2a}\right), \quad \Pi_{1t} \approx \frac{1}{2\pi l^2 a},$$

where a is the cell parameter. Combining (34), (40), and (41), we get

$$Q_l(k) = G_l'^2(k)w\left[\frac{3(\mathbf{k}\cdot\mathbf{P})^2}{k^2} + B_l\left(1 + (4\pi)^2 n_{\rm dis} \times \int \frac{d\mathbf{q}}{(2\pi)^3} R(\mathbf{k},\mathbf{q})G_t'^2(q)\right)\right], \tag{54}$$

$$Q_t(k) \approx G_t'^2(k)\Sigma_t^{12}(0) = G_t'^2(k)wB_l. \tag{55}$$

For $k \gg \theta$ the integral in (54) is almost independent of $\mathbf{k}$,

$$Q_l(k) \approx G_l'^2(k)w\left[\frac{3(\mathbf{k}\cdot\mathbf{P})^2}{k^2} + B_l(1 + 2w\Pi_{2t})\right],$$

while for $k \ll \theta$,

$$Q_l(k) = G_l'^2(k)w\left[\frac{3(\mathbf{k}\cdot\mathbf{P})^2}{k^2} + B_l \times\left(1 + \frac{4\pi n_{\rm dis}k^2(\Delta_l + 4\Delta_t)}{15\theta^{3/2}}\right)\right].$$

Integrating (54) and (55) over $\mathbf{k}$ and using the same line of reasoning as in deriving (49), we arrive at the following equations of state:

$$B_l = w\Pi_{2l}[B_l(1 + 2w\Pi_{2t}) + 3P^2], \tag{56}$$

$$B_t = w\Pi_{2t}B_l. \tag{57}$$

As $\theta \to 0$,

$$\Pi_{2l} \approx \frac{1}{(2\pi)^2 l^3}\left(\frac{1}{2\sqrt{\pi}}\arctan\frac{\sqrt{\pi}l}{2a} - \frac{la}{\pi l^2 + 4a^2}\right),$$

$$\Pi_{2t} \approx \frac{1}{8\pi l^3\sqrt{\theta}}. \tag{58}$$

Let us find the equation of state for P. Substituting $P_{\alpha i} = P$ in (31) and using (37) and (41), we obtain

$$\frac{\partial\Phi(\mathbf{P}_\alpha)}{\partial P_{i\alpha}}=\tau P+vP^3+TP\int\frac{d\mathbf{k}}{(2\pi)^3}[v(G_l^{11}(k)+2G_t^{11}(k))$$
$$-\beta w(G_l^{11}(k)+(n-1)G_t^{12}(k))]=0.$$

Taking n to zero and allowing for (43), we arrive at the equation of state for the polarization P:

$$[\tau'+v(P^2+B_l+2B_t)]P=0. \tag{59}$$

Combining (22), (23), and (33) results in the following expression for the homogeneous dielectric susceptibility $\chi_{ij}=\chi\delta_{ij}$:

$$\chi^{-1}=\tau'+v(B_l+2B_t+3P^2). \tag{60}$$

If we compare (60) with (53), we see that $\chi=\theta^{-1}=G_t'(0)$. For the glassy correlator $S_{ij}=S\delta_{ij}$ Eqs. (24) and (33) yield

$$S=\chi^2wB_l=Q_t(0), \tag{61}$$

while for the contribution to the specific heat we have

$$\delta C=\frac{3T\chi P^2}{T_0^2}+\frac{T^2}{2T_0^2}\left\{\Pi_{2l}+2\Pi_{2t}+2\beta\right.$$
$$\left.\times\int\frac{d\mathbf{k}}{(2\pi)^3}[G_l'(k)Q_l(k)+2G_t'(k)Q_t(k)]\right\}$$
$$\approx\frac{3T\chi P^2}{T_0^2}+\frac{T^2}{2T_0^2}[\Pi_{2l}+2\Pi_{2t}+2\beta w\Pi_{3l}$$
$$\times(B_l+2B_t+3P^2)+4\beta w\Pi_{3t}B_l]. \tag{62}$$

As $\theta\to0$,

$$\Pi_{3l}\approx\frac{1}{(4\pi l)^3}\left[\frac{la(\pi l^2-4a^2)}{\pi l^2-4a^2}+\frac{1}{2\sqrt{\pi}}\arctan\frac{\sqrt{\pi}l}{2a}\right],$$

$$\Pi_{3t}\approx\frac{1}{32\pi l^3\theta^{3/2}}.$$

Combining (16), (35), (41), and (43) at $P_{\alpha i}=P$ and letting n go to zero, we get

$$\Phi=\frac{3}{2}\tau P^2+\frac{3}{4}vP^4-\frac{1}{2}\int\frac{d\mathbf{k}}{(2\pi)^3}[T\ln G_l'(k)$$
$$+2T\ln G_t'(k)+G_l'^{-1}(k)Q_l(k)$$
$$+2G_t'^{-1}(k)Q_t(k)]. \tag{63}$$

If we now plug (52), (54), and (55) into this equation and use (56) and (57), in the lowest order in θ (we discard an unimportant term of the form $-T\times$const) we get

$$\Phi=\frac{3}{2}\tau P^2+\frac{3}{4}vP^4+\frac{1}{2}\theta[T(\Pi_{1l}+2\Pi_{1t})+B_l+4B_t]$$
$$-\frac{1}{2}B_l(\Pi_{1l}\Pi_{2l}^{-1}+2w\Pi_{1t})-\frac{a}{3l}T\Pi_{1t}\theta^{3/2}. \tag{64}$$

The final formulas for the thermodynamic parameters of the crystal can be found by substituting the solutions of Eqs. (56), (57), and (59) for which χ is positive in (60)–(64).

4. PHASE TRANSITIONS

The possible phases in the ferroelectric with dislocations considered here are determined by the possible solutions of the equations of state (56), (57), and (59). The paraelectric phase corresponds to the trivial solution $B_l=B_t=P=0$. In this phase we have $\chi=1/\tau'$ and

$$\Phi_{PE}=\frac{\tau'}{2}T(\Pi_{1l}+2\Pi_{1t})-\frac{a}{3l}T\Pi_{1t}\tau'^{3/2}, \tag{65}$$

and the contribution of the polarization fluctuations to the specific heat is

$$\delta C_{PE}\approx\frac{T^2}{8\pi l^3T_0^2\sqrt{\tau'}}. \tag{66}$$

The phase is stable if $\tau'>0$.

The dipole-glass phase corresponds to a solution with $P=0$ but B_l and B_t finite. For this case Eq. (56) yields

$$1=w\Pi_{2l}(1+22w\Pi_{2t}). \tag{67}$$

Using the assumed smallness of w, in view of which

$$w\Pi_{2l}\ll1, \tag{68}$$

we combine (53), (57), (60), and (67) and get

$$\chi\approx\left(\frac{w^2\Pi_{2l}}{4\pi l^3}\right)^{-2}\equiv\theta_c^{-1}, \tag{69}$$

$$B_l\approx\frac{w\Pi_{2l}(\theta_c-\tau')}{v},\quad B_t\approx\frac{B_l}{2w\Pi_{2l}}. \tag{70}$$

Thus, for $\tau'<\theta_c$ there is a positive solution for B_l, with the result that we have positive Edwards–Anderson parameters $Q_l(k)$ and $Q_t(k)$, i.e., there is a transition to the glassy phase. The thermodynamic potential in this phase is

$$\Phi_{DG}\approx\frac{1}{2}\theta_cT(\Pi_{1l}+2\Pi_{1t})-\frac{w\Pi_{1l}}{2v}(\theta_c-\tau')$$
$$-\frac{a}{3l}T\Pi_{1t}\theta_c^{3/2}, \tag{71}$$

and the contribution to the specific heat is

$$\delta C_{DG}\approx\frac{T^2}{8\pi l^3T_0^2\sqrt{\theta_c}}\left(1+\frac{w^2\Pi_{2l}(\theta_c-\tau')}{4vT\theta_c}\right). \tag{72}$$

The glassy phase is stable in the entire range of its existence, $\tau'<\theta_c$.

Before we consider the transition from the paraelectric phase to the glassy phase, let us discuss the limits within which the above results are valid. The mean-field region within which the results are valid is determined by the condition for the smallness of the fluctuation contribution to the renormalization of the coupling constant v (see Ref. 12), $vT_c\Pi_{2t}\ll1$, or, assuming that $l\sim a$,

$$vT_c\Pi_{1s}\ll\sqrt{\theta},\quad s=l,t. \tag{73}$$

This is the only condition because $F_{\rm dis}$ of Eq. (12) (and the coupling constant w) is not renormalized in the region of large fluctuations, since it contains two longitudinal Fourier

transforms of the polarization with a finite susceptibility at the transition point [see the first equation in (52)]. The condition (73) implies that for the region of interest to us, $\theta \sim \theta_c$, to belong to the mean-field region, the following condition must be met:

$$v T_c \ll w. \tag{74}$$

In this case $\Phi_{DG} < \Phi_{PE}$ for all $\tau' < \theta_c$, so that the phase transition to the glassy phase is a second-order transition at point $\tau' = \theta_c$.

Let us now consider the ferroelectric (rhombohedral) phase, which corresponds to a solution with P, B_l, and B_t finite. This phase meets the following conditions:

$$\theta = 2vP^2, \tag{75}$$

$$B_l = \frac{w\Pi_{2l}(\theta - \tau')}{v}, \tag{76}$$

and the reciprocal susceptibility θ satisfies the equation

$$\theta^{3/2} + 2\theta_c^{1/2}\theta + 2\tau'(\theta^{1/2} - \theta_c^{1/2}) = 0. \tag{77}$$

For $\tau' < -\mu\theta_c \equiv \tau_0$, $\mu = (\sqrt{3993} + 59)/16 \approx 7.637$, this equation has two solutions satisfying the condition $\theta > \tau'$, which ensures that B_l is positive [see the condition (76)]. Near τ_0 these solutions have the form

$$\theta_\pm = \theta_0 \pm \sigma\sqrt{\theta_0(\tau_0 - \tau')}, \quad \theta_0 \equiv v\theta_c,$$

$$v = \frac{4}{9}\left[\left(1 + \frac{45}{8}\mu\right)^{1/3} - 1\right]^2 \approx 2.782,$$

$$\sigma = \left(1 + \frac{45}{8}\mu\right)^{-2/3}\sqrt{2(\mu^2 - 2)} \approx 0.866.$$

For $\tau' \ll -\theta_c$ the larger of these two solutions, θ_+, tends to $-2\tau'$ and the smaller, θ_-, to θ_c. With allowance for the condition (74), the potential of the ferroelectric phase has the form

$$4v\Phi_{FE} \approx (\theta + 2\tau')(w\Pi_{1l} - 4w^2\Pi_{2l}\Pi_{1t}) + \tau'(3\theta - 4\sqrt{\theta_c\theta}) + 4\sqrt{\theta_c\theta^3} + \frac{3}{4}\theta^2.$$

Comparing this at $\theta = \theta_\pm$ with Φ_{DG}, we see that at $\tau' \approx -2w\Pi_{1l} \equiv \tau_p$ there is a second-order phase transition to the ferroelectric phase corresponding to the solution θ_+. For $\tau' < \tau_p$ we have $\theta_+ \approx -2\tau' = \chi^{-1}$, so that

$$P = \sqrt{-\frac{\tau'}{v}}, \quad B_l = \frac{w}{v}\Pi_{2l}(-3\tau'),$$

$$B_t = \frac{3}{2v}\sqrt{-\frac{\theta_c\tau'}{2}},$$

$$\delta C \approx \frac{3T}{2vT_0^2} + \frac{T^2}{T_0^2 8\pi l^3\sqrt{-2\tau'}}\left(1 + \frac{3w^2\Pi_{2l}}{3vT}\right).$$

The temperature dependence of the obtained thermodynamic quantities near glassy and ferroelectric transitions is depicted in Fig. 1.

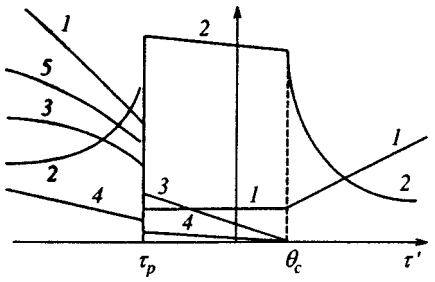

FIG. 1. Temperature dependence of the thermodynamic quantities for a ferroelectric with dislocations near glassy and ferroelectric phase transitions: *1*, the reciprocal dielectric susceptibility $\theta = \chi^{-1}$; *2*, the specific heat; *3*, B_t; *4*, B_l; and *5*, polarization P.

5. CONCLUSION

Let us discuss some qualitative features of the dipole-glass phase in a ferroelectric with dislocations and the possibility of studying this phase in experiments. From the theoretical viewpoint, the most important feature of this phase is the occurrence of the Edwards–Anderson nonlocal parameter $Q_{ij}(\mathbf{k}) = \langle\langle P_i(\mathbf{k})\rangle_T\langle P_j(-\mathbf{k})\rangle_T\rangle_C$. In the existing models of glassy transitions, $Q_{ij}(\mathbf{k})$ is independent of $\mathbf{k}$ (see Ref. 8), which means that there are no correlations between the spontaneous local dipole moments that appear in the glassy phase: $Q_{ij}(\mathbf{r}) = Q_{ij}\delta(\mathbf{r})$. In our case, however, such correlations are present and extend over distance of order of the longitudinal and transverse correlation radiuses [see Eqs. (54) and (55)], $\xi_t = l/\sqrt{\theta}$ and $\xi_l = l/\sqrt{\theta + 4\pi}$. Since $\xi_t \gg a$, we can expect that in the glassy phase there may be regions of correlated vortex (transverse) polarization whose size is of order ξ_t. The existence of such regions should manifest itself, in particular, in the diffuse scattering of x rays and neutrons, which is described by the inhomogeneous susceptibility $G_{ij}(\mathbf{k})$. According to (42) and (43), in the glassy phase there appears a contribution to $G_t(k) = G_t'(k) + Q_t(k)$ proportional to the square of a Lorentzian with a halfwidth ξ_t^{-1}. Experimental observations of such an effect will make it possible to establish the point of transition to the dipole-glass phase and to determine the temperature dependence of the quantity B_l, which acts as the order parameter for this transition. Note that instead of B_l we can consider the common Edwards–Anderson order parameter

$$Q = \int \frac{d\mathbf{k}}{(2\pi)^3}\,\mathrm{Tr}\,\hat{Q}(\mathbf{k}) = B_l + 2B_t = B_l(1 + 2w\Pi_{2t}).$$

Generally speaking, the replica-symmetric description of the glassy phase by a single order parameter may prove to be inadequate in view of replica instability.[8] In this paper we did not check solutions for replica stability, which in the model considered here may prove to be a complicated task. Our aim was only to demonstrate the possibility, in principle, of the existence of an intermediate glassy phase in a ferroelectric with dislocations. In view of this, the temperature dependences of the susceptibility obtained in this paper are of a preliminary qualitative nature and demonstrate the qualitative features to which the existence of an intermediate glassy phase, in particular, may smooth out dielectric anomalies.

It must also be noted that, in addition to the spherically symmetric solutions for $G_s(\mathbf{k})$ and $Q_s(\mathbf{k})$ considered in this paper, there are solutions with an angular dependence of these quantities, determined by one of the spherical harmonics Y_{lm}. Generally speaking, due to the additional inhomogeneity, such solutions have a higher thermodynamic potential and may be metastable, but a thorough study of this problems merits a separate (and very complicated) investigation.

This work was sponsored by the Russian Fund for Fundamental Research (Grants Nos. 98-02-18069 and 97-02-17878).

*)E-mail: timonin@icomm.ru

[1] A. B. Harris and T. C. Lubensky, Phys. Rev. Lett. **33**, 1540 (1974).
[2] D. E. Khmel'nitskiĭ, Zh. Éksp. Teor. Fiz. **68**, 1960 (1975) [Sov. Phys. JETP **41**, 981 (1975)].
[3] S. L. Ginzburg, Zh. Éksp. Teor. Fiz. **73**, 1961 (1977) [Sov. Phys. JETP **46**, 1029 (1977)].
[4] S. N. Dorogovtsev, Zh. Éksp. Teor. Fiz. **80**, 2053 (1981) [Sov. Phys. JETP **53**, 1070 (1981)].
[5] A. L. Korzhenevskiĭ, Fiz. Tverd. Tela (Leningrad) **28**, 1324 (1986) [Sov. Phys. Solid State **28**, 745 (1986)].
[6] V. M. Nabutovskiĭ and B. Ya. Shapiro, Zh. Éksp. Teor. Fiz. **75**, 948 (1978) [Sov. Phys. JETP **48**, 480 (1978)].
[7] I. M. Dubrovskiĭ and M. A. Krivoglaz, Zh. Éksp. Teor. Fiz. **77**, 1017 (1979) [Sov. Phys. JETP **50**, 512 (1979)].
[8] K. Binder and A. P. Young, Rev. Mod. Phys. **58**, 801 (1986).
[9] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, New York (1984), Chap. 2.
[10] P. N. Timonin, Zh. Éksp. Teor. Fiz. **110**, 914 (1996) [JETP **83**, 503 (1996)].
[11] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, Pergamon Press, New York (1970), Chap. 4.
[12] A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions*, Nauka, Moscow (1982), Chap. 10 [English translation of an earlier edition: Pergamon Press, Oxford (1979)].

Translated by Eugene Yankovsky

Mössbauer spectra of Stoner–Wohlfarth particles in rf fields in a modified relaxation model

A. M. Afanas'ev and M. A. Chuev

Physical Engineering Institute, Russian Academy of Sciences, 117218 Moscow, Russia

J. Hesse

Institut für Metallphysik und Nukleare Festkörperphysik, Technische Universität Braunschweig, D-38106 Braunschweig, Germany

(Submitted 6 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1001–1026 (September 1999)

A theory of Mössbauer absorption spectra in the presence of external rf fields is developed for Stoner–Wohlfarth particles in an extended relaxation model with a more realistic description of the relaxation process. Calculating in this model, we track the transformation of the Mössbauer spectra as a function of relaxation processes for the transitional region and frequencies and amplitudes of the rf field where the well-resolved hyperfine structure transitions into an isolated central peak with satellites. In this transitional region new types of resonance effects are found which have no analogy with previously known resonance phenomena. © *1999 American Institute of Physics.* [S1063-7761(99)01709-6]

1. INTRODUCTION

At present, a large amount of experimental material has accumulated demonstrating a wide variety in the shape of the transformation of Mössbauer absorption spectra as a function of the frequency and amplitude of the rf field in magnetic systems consisting of magnetic clusters of small (nanometer) dimensions.[1–9] In Ref. 10 we developed a theory of Mössbauer absorption spectra in the presence of an rf magnetic field for the case of an ensemble of Stoner–Wohlfarth (SW) particles which included relaxation processes in full agreement with the original model (Ref. 11). The theory we developed allows one to calculate Mössbauer spectra for arbitrary values of the amplitude and frequency of the rf field. An important consequence of this theory is the presence of a sharp transition from the well-resolved hyperfine structure for small amplitudes of the rf field to an isolated central line with satellites for large amplitudes of the rf field. This prediction of the theory does not find confirmation in existing experimental studies, primarily because of the simplified description of the relaxation process in the original SW model.[11]

The present paper generalizes the results of preceding studies to the case of a broader description of the relaxation process in a system of SW particles. In the initial SW model the time dependence of the hyperfine field at the nucleus can be described by a strictly deterministic trajectory, i.e., at all times the magnitude of the hyperfine field at the nucleus for any particle, depending on the frequency and amplitude of the rf field, is exactly determined. In the extended SW relaxation model proposed here, the trajectories acquire a stochastic character, so that for each individual particle one cannot indicate exactly the magnitude and direction of the magnetic moment (and, consequently, the the hyperfine field at the nucleus) of the particle at every instant of time. It is possible only to determine the probabilities of finding a particle in some quasi-equilibrium state corresponding to various energy minima. This latter fact greatly complicates the problem, especially its numerical realization.

On the other hand, even small deviations of the relaxation process from the original SW model give rise to noticeable changes in the Mössbauer spectra in the region of the transition from the allowed hyperfine structure to an isolated central line and, moreover, to new physical phenomena. In this transitional region qualitatively new types of resonances are realized at frequencies of the rf field coupled by the parametric resonance conditions with the frequencies of the components of the magnetic hyperfine structure. These resonances differ, both in their position and in the manner in which they are manifested, from previously known resonances at frequencies corresponding to splitting of the energy levels in the ground state and excited states of the nucleus.[10,12,13]

Section 2 formulates the modified SW model with a more general description of the relaxation process. Section 3 obtains a general expression for the Mössbauer absorption spectra in the extended SW relaxation model on the basis of the approaches proposed earlier in Refs. 10 and 14. The use of general formulas to describe the Mössbauer spectra requires, generally speaking, complicated computer calculations, which necessitates a deeper approach to the analysis of the calculational methods prior to the writing of the computer program. These questions are considered in Sec. 4, which also gives examples of calculations of Mössbauer spectra using the described technique.

On the basis of the numerical calculations, Sec. 5 analyzes the transformation of the Mössbauer spectra as a function of the parameters of the relaxation model for the region of transitional amplitudes of the rf field, where the resolved hyperfine structure goes over to an isolated central line with satellites. It is found that in a narrow interval of

frequencies of the rf field near multiples of the frequencies of the components of the magnetic hyperfine structure, depending on the multiplicity of the indicated frequencies, the lines undergo both a relaxational narrowing and a mutual repulsion. A simplified model is proposed to reveal the physical essence of these phenomena.

2. THE GENERALIZED STONER–WOHLFARTH MODEL

To describe the remagnetization processes in magnetic alloys, wide use is made of the model proposed in 1948 by Stoner and Wohlfarth.[11] This model considers a ferromagnet as a set of individual particles or clusters, inside which a strong exchange interaction takes place, so that each particle can be treated as being uniformly magnetized with magnetization M_0. It is assumed here that each particle has the shape of a prolate ellipsoid of revolution with the axes of these particles randomly distributed in space, so that the total Hamiltonian of such a particle in an external magnetic field **H** can be represented as

$$\hat{H} = -HVM_0 \cos\phi - KV\cos^2(\theta-\phi), \tag{1}$$

where V is the volume of the particle, K is the magnetic anisotropy constant, θ is the angle between the direction of the external field and the easy axis of the particle, and ϕ is the angle between the direction of the magnetic moment and the external field.

The SW model also assumes that

$$KV \gg k_B T, \tag{2}$$

so that in the absence of external fields the magnetic moment of each particle is either aligned or anti-aligned with the easy axis. As a consequence of relation (2) hops between these two states are assumed to be very slow, so that during the measurement time they can be neglected.

When a magnetic field is applied the energy minima with respect to the angle ϕ are shifted, i.e., the magnetic moment is deflected from the easy axis (with the exception of particles with $\theta=0$). For weak magnetic fields, two energy minima separated by a maximum are preserved.

If the external field, normalized to the anisotropy energy

$$h = \frac{HM_0}{2K} \tag{3}$$

exceeds some critical value h_c depending on the orientation of the particle,[11] one of the minima disappears, and we have curves with one minimum. Figure 1 plots the dependence of the critical field on the orientation of the particle. As can be seen from the figure, h_c has a minimum at $\theta=45°$ and a maximum at the angles $\theta=0$ and 90°. An important consequence of this model is that even a system of noninteracting SW particles possesses pronounced nonlinear magnetic properties, specifically, the presence of a hysteresis loop, and also such hysteresis-generated characteristics as residual magnetization, a critical field, etc. As will be shown below, this property of an ensemble of SW particles in and of itself defines the specific dependence of the shape of the Mössbauer spectra on the amplitude of the external rf field. In reality the critical fields for the majority of particles lie in the interval 0.5–0.75 (see Fig. 1).

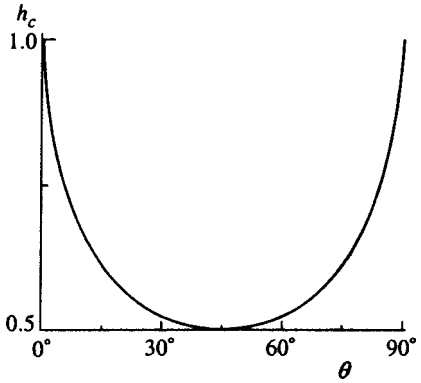

FIG. 1. Dependence of the critical field h_c of a SW particle on the angle θ between the direction of the external magnetic field and the easy axis in the original SW model.

According to the SW model the dynamics of such an ensemble is the following. If the external field exceeds h_c in magnitude, then the particle has only one energy minimum and is always found at this minimum. As the field is decreased, say, from positive values, the magnetic moment of the particle tracks the position of the local minimum and instantaneously adjusts its direction in accordance with it, and only for fields $h<-h_c$ does the magnetic moment instantaneously transition to its absolute minimum. When the amplitude of the external periodic magnetic field is greater than h_c, the time dependence of the magnetization of the particle has the form of a hysteresis loop. If the amplitude of the periodic field is less than h_c, then according to the SW model the particle is found in one of two minima and never leaves it. Curves of the corresponding dependence of the magnetization of the particle on the intensity of the external periodic magnetic field

$$H(t) = H_0 \sin(\omega_{rf} t) \tag{4}$$

are plotted in Fig. 2, which plots projections of the magnetic

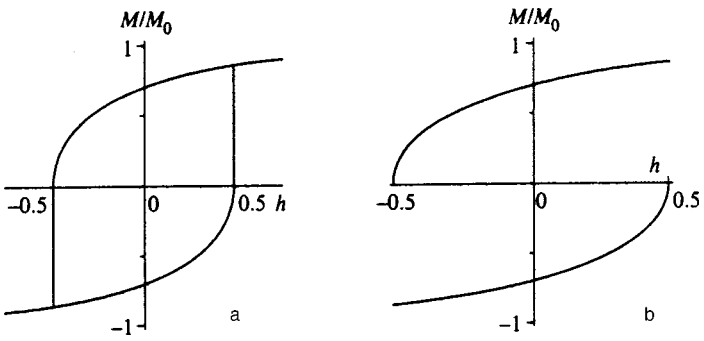

FIG. 2. Magnetization curves of a group of SW particles with orientation $\theta=45°$ for amplitudes of the periodic magnetic field $h_0=0.75$ (a) and 0.5 (b).

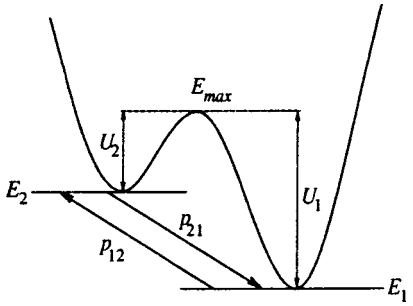

FIG. 3. Scheme of transitions between local energy minima in the generalized SW relaxation model.

moment onto the direction of the field for a group of particles with orientation $\theta=45°$ for different amplitudes of the periodic field h_0, normalized to the magnetic anisotropy constant [see Eq. (3)]. Naturally, the magnetization curve for an ensemble of SW particles will be the result of averaging both over particles with different orientations and over different trajectories within one group in those cases where the amplitude of the external periodic field does not exceed its critical value for this group. For amplitudes $h_0<0.5$ the curve has the form of the reversible magnetization curve. In the interval $0.5<h_0<1$ the magnetization curve acquires the form of loops; however, in essence it is a superposition of hysteresis loops and open curves of the type depicted in Figs. 2a and 2b (see Ref. 10).

A system of SW particles possesses a number of characteristic features, foremost of which is the fact that its magnetization curves, including the hysteresis loops, do not depend on the frequency of the periodic magnetic field. The curves depicted in Fig. 2 should not vary with ω_{rf} if the system is described exactly by the SW model.

Although the relaxation process is nowhere considered directly in Ref. 11, relaxation is nevertheless present implicitly although it enters in a somewhat nontrivial form. Specifically, it is assumed that as the magnitude of the field varies, the magnetic moment of a particle located at a definite energy minimum will, on the one hand, track the variation of the position of this minimum and alter its direction in accordance with the magnitude of the applied field and that, on the other hand, transitions between states with different energy minima are assumed to be forbidden until the magnitude of the applied field exceeds its critical value. In other words, the relaxation process is simultaneously limitingly fast and limitingly slow. The impossibility of hops between states with different energy minima is implicitly due to the fact that there are high energy barriers hindering such transitions (see Fig. 3). This assumption turns out to be valid only in magnetic fields substantially less than h_c; near the critical field it should break down. It is not hard to show that near h_c the height of the barrier is given by

$$U_{2i}(h)=\alpha_i U_0(h_c-|h|)^2, \quad |h|<h_c, \tag{5}$$

where α_i is some numerical constant which depends on the orientation of the ith particle, and $U_0=KV$ is the height of the energy barrier for $H=0$.

For amplitudes of the periodic field near h_c, the particle will be found a significant part of the time in a situation where the assumption of a large value of the barrier is clearly not fulfilled [see Eq. (4)]. As a consequence, a more accurate description of the relaxation process is needed. Specifically, we assume that a sudden reorientation of the magnetic moment of the particle will occur not only for magnetic fields exceeding h_c, but also for weaker fields, where the effective anisotropy energy U_{2i} is not too large relative to the temperature. As we will see below, such a generalization of the SW model radically alters both the magnetic properties and the Mössbauer spectra of the SW particles.

For simplicity we assume that at any instant of time, depending on the magnitude of the magnetic field, the relaxation process is governed by only two quantities:

$$p_{12}(h)=p_0\exp[-U_1(h)/k_BT],$$
$$p_{21}(h)=p_0\exp[-U_2(h)/k_BT], \tag{6}$$

where p_0 is some constant and

$$U_{1,2}(h)=E_{\max}(h)-E_{1,2}(h). \tag{7}$$

Here p_{12} and p_{21} have the meaning of transition probabilities per unit time between local-equilibrium levels. For each group of particles with different orientations the quantities $E_{\max}$ and $E_{1,2}$ are found by fairly straightforward numerical calculations.[11] The proposed relaxation model possesses the virtue that the entire relaxation process is characterized by only two parameters, the constant p_0 and the barrier height U_0. For fixed p_0 and limitingly large U_0 the original SW model is realized, and as U_0 is increased the time interval where the barriers cannot be taken to be small decreases and tends to zero in the limit $U_0\to\infty$.

In magnetic field $|h|<h_c$ each particle can be found only in two states corresponding to local energy minima, between which relaxation hops can occur. Their equilibrium populations $w_{1,2}^{(0)}(h)$ are determined by the principle of detailed balance

$$w_{1,2}^{(0)}(h)=\frac{\exp[-E_{1,2}(h)/k_BT]}{\exp[-E_1(h)/k_BT]+\exp[-E_2(h)/k_BT]}. \tag{8}$$

Naturally, in the presence of an external variable magnetic field the true populations of the local states will not be in equilibrium but will depend on the amplitude of the field and the relation between its frequency and the relaxation parameters. Thus, at all times the variation of the nonequilibrium populations of the local states $w_1(t)$ and $w_2(t)$ in time can be described by the equations

$$\frac{dw_{1,2}(t)}{dt}=\mp p_{12}(t)w_1(t)\pm p_{21}(t)w_2(t). \tag{9}$$

In magnetic fields exceeding the critical field $|h|>h_c$, according to the original SW model there are no longer two, but rather only one local minimum of the energy of the particle, and according to this same model the particle undergoes a transition to the state corresponding to this minimum in magnetic fields exceeding h_c by an arbitrarily small amount. In what follows it will be convenient in such fields to maintain two states of the particle with identical values of the magnetic moment and to assume that at times corresponding to $|h|>h_c$ the true populations do not depend on time and that the population vector $\langle W(t)|$ has the form

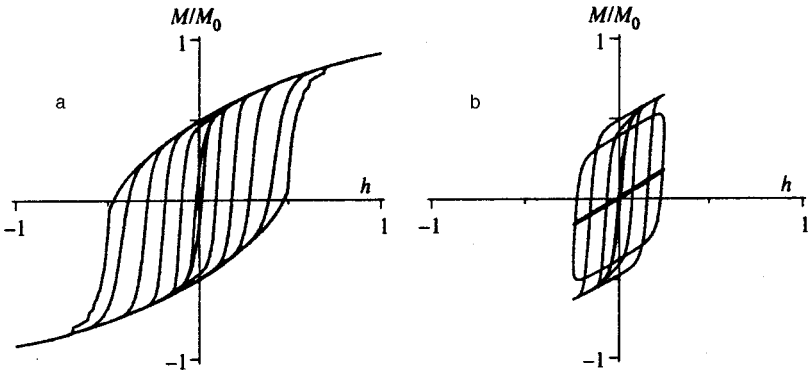

FIG. 4. Magnetization curves for an ensemble of randomly oriented SW particles in the modified relaxation model for $U_0/k_BT=20$ and assorted values of the ratio $p_0/(\omega_{rf}/2\pi)=1,\ 10^2,\ 10^4,\ 10^6,\ 10^8,\ 10^{10}$ (from the outer hysteresis loop inward) in the regime of strong rf fields, $h_0=1$ (a) and in the regime of weak rf fields, $h_0=0.25$ (b).

$$\langle W(t)|=(w_1(t)w_2(t))=\begin{cases}(1\ \ 0), & h>h_c,\\ (0\ \ 1), & h<-h_c.\end{cases} \tag{10}$$

For amplitudes of the external magnetic field exceeding the critical field, condition (10) plays the role of a boundary condition ensuring uniqueness of the solution of Eq. (9). For fields with amplitude $h_0<h_c$, we choose the condition of periodicity

$$\langle W(t+T_{rf})|=\langle W(t)|, \tag{11}$$

as the boundary condition, where $T_{rf}=2\pi/\omega_{rf}$ is the period of the external magnetic field.

The system of nonlinear equations (9) together with the boundary conditions (10) and (11) is relatively easy to solve numerically. Knowing $w_1(t)$ and $w_2(t)$, it is not hard to determine the evolution of the magnetic moment of the particle in time under the action of the external magnetic field:

$$M(t)=w_1(t)M_1(h(t))+w_2(t)M_2(h(t)), \tag{12}$$

where $M_1(h)$ and $M_2(h)$ are the projections, corresponding to the local energy minima, of the magnetic moment of the particle onto the direction of the external field, which are calculated independently according to the scheme set forth in Ref. 11.

To determine the evolution in time of the magnetic moment $\mathbf{M}(t)$ of an ensemble of randomly oriented, noninteracting SW particles, it is necessary to sum up the magnetic moments of particles of the ith type $\mathbf{M}_i(t)$, each of which is determined by the set of equations (6)–(11).

Generalization of the relaxation process produces dramatic qualitative changes in the nonlinear magnetic properties of the system, primarily to changes in the shape of the hysteresis loops. In the original SW model the shape of the hysteresis loop is determined only by the amplitude of the variable magnetic field and does not depend on its frequency, whereas in the generalized relaxation model it depends on the frequency of the field, specifically on the ratio ω_{rf}/p_0, and also on the relative magnitude of the energy barrier U_0/k_BT.

Figure 4a shows hysteresis loops as a function of the ratio ω_{rf}/p_0 in the regime of large amplitudes of the rf field h_0 exceeding the values of the critical fields for all particles with different orientations. As can be seen from the figure, as the frequency of the magnetic field increases the hysteresis loop broadens and in the limit of large ω_{rf} tends to the hysteresis loop characteristic of the original SW model, and the remagnetization field, naturally, grows as the frequency of the field is increased. This picture has a simple physical explanation. At low frequencies the particles take a long time to thermalize and reach equilibrium, and the remagnetization fields are small. At high frequencies ω_{rf} the magnetic field does not have enough time to repopulate the local energy levels in accordance with their equilibrium populations. In this case, remagnetization requires larger amplitudes of the rf field.

Note that this fact may have a direct bearing on the Mössbauer experiments performed in the pioneering work of Pfeiffer on the collapse effect,[1] in which the resolved hyperfine structure ''collapses'' into an isolated central line with satellites as the amplitude of the rf field is increased. Pfeiffer worked with permalloy and a frequency of the external rf field on the order of 100 MHz, where the collapse effect and, consequently, effective remagnetization of the magnetic clusters occur in rf fields of order several oersteds. On the other hand, permalloy is a magnetically soft material, and it remagnetizes at low frequencies even in fields on the order of hundredths of an oersted. Pfeiffer did not draw attention to this fact. Moreover, in his subsequent works he did not compare the remagnetization fields at high and low frequencies of the rf field. On the other hand, it is clear that to construct a consistent theory of Mössbauer spectra it is necessary to take this fact into account. The generalized relaxation model proposed here makes it possible to describe this phenomenon.

In weak rf fields with amplitudes less than critical (h_c), for all particles (see Fig. 4b) the shape of the hysteresis loops as a function of the frequency ω_{rf} varies in a more complicated way. In the low-frequency region, growth of ω_{rf} the remagnetization fields strengthen. In the limit of high frequencies, where the relaxation process becomes unimportant, the hysteresis loops disappear and a reversible magnetization curve of paramagnetic type appears. In the region of intermediate frequencies ω_{rf} exotic loops of the type depicted in Fig. 4b are observed. Of course, in this case one can expect the appearance of special effects in the Mössbauer spectra.

3. RELAXATION MÖSSBAUER SPECTRA OF STONER–WOHLFARTH PARTICLES: BASIC FORMULAS

The rf field affects the hyperfine structure spectra via the hyperfine field $\mathbf{H}_{hf}(t)$ acting on the nucleus, where this latter field, acted upon by the rf field via complex relaxation pro-

cesses, begins to depend on time. The simplest and physically clearest situation is realized in the case where the relaxation processes are quite fast, so that $\mathbf{H}_{hf}(t)$ tracks some of the macroscopic characteristics of the sample, e.g., the magnetization $\mathbf{M}_i(t)$ of the SW particle, which in turn varies in time under the action of the rf field. If the magnetization model is prescribed, i.e., if $\mathbf{M}_i(t)$ and, consequently, $\mathbf{H}_{hf}(t)$ can be assumed to be known, then the Hamiltonian of the system will have the form

$$\hat{H}=\hat{H}_0+g_{g,e}\mu_N\hat{\mathbf{I}}^{(g,e)}\mathbf{H}_{hf}(t)+\hat{V}_{\gamma N}(t_0), \tag{13}$$

where $\hat{H}_0$ is the Hamiltonian determining the energy levels of the nucleus without allowance for the hyperfine interaction; the second term describes the hyperfine interaction of the nucleus, μ_N is the nuclear magneton, $g_{g,e}$ is the nuclear g factor, $\hat{\mathbf{I}}^{(g,e)}$ is the nuclear spin for the ground (g) state and the excited (e) states of the nucleus; the operator $\hat{V}_{\gamma N}(t_0)$ describes the interaction of a γ quantum with the nucleus, and t_0 is the time at which this interaction is switched on.[15] As can be seen from Eq. (13), the shape of the absorption spectrum is determined by $\mathbf{H}_{hf}(t)$.

In Ref. 10, where the time dependence of $\mathbf{H}_{hf}$ was assumed to be strictly determined in accordance with the original SW model, we obtained a general expression for the absorption cross section of a γ quantum of energy $E=\hbar\omega$ for a group of SW particles with prescribed orientation of the easy axis

$$\sigma_i(\omega)=\frac{\sigma_\alpha}{\Gamma_0 T_{rf}}\int_0^{T_{rf}}dt\int_t^\infty dt'\times\sum_\eta \mathrm{Tr}\left\{\hat{V}_\eta\left[\hat{T}\exp\left\{\int_t^{t'} i[\tilde{\omega}\hat{\mathbf{1}}_n-\hat{\mathbf{L}}_{\hat{H}}(t'')]dt''\right\}\right]\hat{V}_\eta^+\right\}+\text{c.c.}, \tag{14}$$

where $\hat{T}$ is the time-ordering parameter, $\hat{\mathbf{L}}_{\hat{H}}(t)$ is the Liouville superoperator of the static hyperfine interaction,[16] acting in a space of $(2I^{(g)}+1)(2I^{(e)}+1)$ variables (the explicit form of this operator for the case of ^{57}Fe nuclei is given in the Appendix), $\hat{\mathbf{1}}_n$ is the unit superoperator in the space of the nuclear variables, $\hat{V}_\eta(t)$ is the nuclear current operator, which defines the intensity of the nuclear transitions, σ_a is the effective thickness of the absorber, and $\tilde{\omega}=\omega+i\Gamma_0/2$ (Γ_0 is the width of the excited energy level of the nucleus). The sum in formula (14) is over polarizations η of the incident γ radiation.

Figure 5 (left) shows an example of the transformation of Mössbauer spectra of an ensemble of SW particles with variation of the amplitude of the rf field, calculated in the original SW model. A characteristic feature of the spectra in their transformation from a resolved hyperfine structure to an isolated line with satellites is the presence of an abrupt qualitative change in the spectra at amplitudes of the rf field near the lower critical field $h_c=0.5$. Even a small (on the order of a few percent) excess of the amplitude of the rf field above its critical value $h_c=0.5$ gives rise to a distinct central line with small satellites superposed on the well-resolved hyperfine structure.

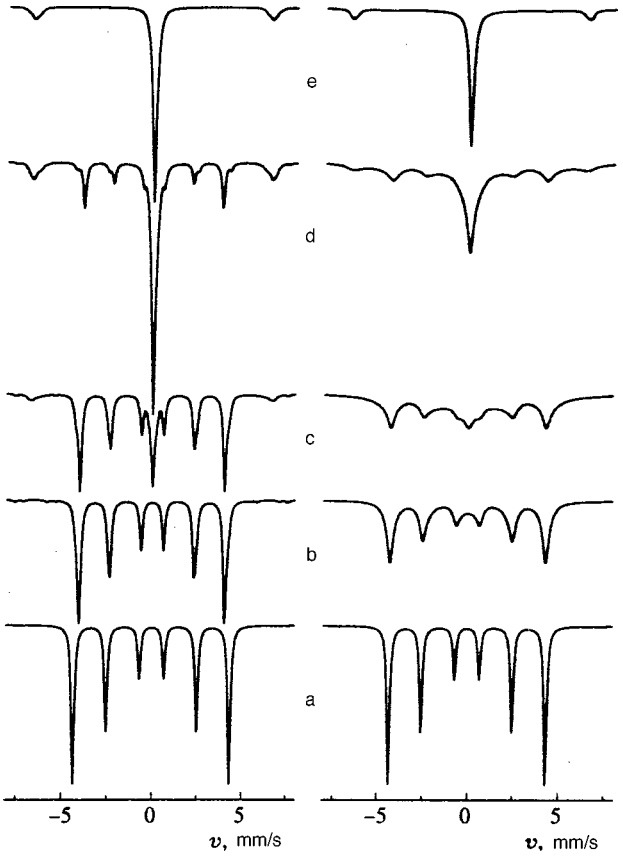

FIG. 5. Transformation of the Mössbauer spectra of an ensemble of randomly oriented SW particles with variation of the amplitude of the rf field ($\omega_{rf}/2\pi=75$ MHz) in the original SW model (left) for $h_0=0$, 0.5, 0.51, 0.6, 1 (curves a–e) and in the extended relaxation model for $U_0/k_BT=20$, $p_0/(\omega_{rf}/2\pi)=10^4$ (right) for $h_0=0$, 0.2, 0.225, 0.25, 0.3 (curves a–e). The scale of spectra e is reduced by a factor of two. Here and below, the calculations are for ^{57}Fe nuclei, $\omega_L/2\pi=50$ MHz, and an unpolarized beam of γ quanta, perpendicular to the direction of the external magnetic field.

In the extended SW relaxation model considered in Sec. 2 the time trajectories of $\mathbf{H}_{hf}$ already acquire a stochastic character, i.e., for each individual particle one cannot exactly indicate the magnitude of the magnetic moment of the particle (and consequently the hyperfine field at the nucleus) at every instant of time. In this case it is possible only to define the probabilities of finding the particle in some quasi-equilibrium state corresponding to different energy minima of the particle. As a consequence, expression (14) must be averaged over all possible trajectories of $\mathbf{H}_{hf}(t)$.

Let us start with the case where the amplitude of the external rf field for all of the SW particles is less than critical:

$$h_0<h_c. \tag{15}$$

In this case, at every instant of time each particle has two distinct energy minima. The original SW model does not assume any transitions between these two states, as if by virtue of a large energy barrier separating these states. In reality, the magnitude of this barrier is finite, and in all external fields near critical it can be as small as desired and, consequently, quite rapid relaxation processes can arise. Here, the transition probabilities between local levels per unit time [see expressions (6)] completely determine the entire set of stochastic curves. Averaging over these stochastic

curves is performed in accordance with the procedure proposed long ago by Anderson[17] for the case where relaxation is governed by a uniform Markov process with constant (in time) transition probabilities. In Ref. 14 this approach was generalized to the case where the transition probabilities depend on time, assuming that the hyperfine field at the nucleus varies only in magnitude while remaining parallel to one of the axes. This latter assumption allows us to avoid having to introduce superoperators of the hyperfine interaction into the theory, which must be introduced when not only the magnitude but also the direction of the hyperfine field varies in the dynamics of the hyperfine fields. Here, an account of the latter effect even without generalizing the relaxation model yields qualitatively different transformations of the Mössbauer spectra as a function of the frequency and amplitude of the rf field.[10] For a more complete description of all possible variants of the effect of an rf field on the Mössbauer spectra, it is of course necessary to consider a model in which, on the one hand, relaxation processes are taken into account and, on the other, the hyperfine field can vary its direction in time.

If condition (15) is satisfied, then generalization of the results of Refs. 10 and 14 by going over to superoperators of a more general form will allow us to obtain the following expression for the absorption cross section in the generalized SW relaxation model:

$$\sigma_i(\omega)=\frac{\sigma_\alpha}{\Gamma_0 T_{rf}}\int_0^{T_{rf}}dt\int_t^\infty dt'\langle W_i(t)|\sum_\eta \mathrm{Tr}\{\hat V_\eta \hat{\mathbf{G}}_i(t,t') \times\exp[i\tilde\omega(t'-t)]\hat V_\eta^+\}|1\rangle+\text{c.c.}, \tag{16}$$

where

$$\hat{\mathbf{G}}_i(t,t')=\hat T\exp\left\{\int_t^{t'}dt''[-i\hat{\mathbf{L}}(t'')-\hat{\mathbf{P}}(t'')]\right\}. \tag{17}$$

Here

$$\hat{\mathbf{L}}(t)=\begin{pmatrix}\hat{\mathbf{L}}_{\hat H}^{(1)}(t) & 0\\ 0 & \hat{\mathbf{L}}_{\hat H}^{(2)}(t)\end{pmatrix}, \tag{18}$$

$$\hat{\mathbf{P}}(t)=\hat P(t)\otimes\hat{\mathbf{1}}_n, \tag{19}$$

$$\hat P(t)=\begin{pmatrix}p_{12}(t) & -p_{12}(t)\\ -p_{21}(t) & p_{21}(t)\end{pmatrix}, \tag{20}$$

where the symbol $\otimes$ stands for the direct product of the operators. In our case, the superoperator $\hat{\mathbf{G}}_i(t,t')$ acts in a space of variables with dimension

$$N=2(2I^{(g)}+1)(2I^{(e)}+1). \tag{21}$$

Here the superoperators of the hyperfine interaction $\hat{\mathbf{L}}_{\hat H}^{(j)}(t)$ act only in the space of nuclear variables and correspond to two different local minima of the energy of the SW particle in the external magnetic field, and the relaxation matrix $\hat{\mathbf{P}}(t)$ is diagonal in the nuclear variables and is defined by the relaxation parameters $p_{12}(t)$ and $p_{21}(t)$ introduced above. Of course, the matrices $\hat{\mathbf{L}}(t)$ and $\hat{\mathbf{P}}(t)$ are different for particles with different orientations, and the particle index i has been dropped for simplicity. Note that the superoperators $\hat{\mathbf{L}}(t)$ and $\hat{\mathbf{P}}(t)$ are completely determined by the characteristics of the original SW model and the characteristics of the rf field; however, on the whole the Mössbauer spectra of the particles are still determined by the dynamics of the system of particles via the vector of nonequilibrium populations $\langle W_i(t)|$ appearing in Eq. (16), to find the components of which it is necessary to solve system of equations (9)–(11) for each type of particle.

In complete analogy with the results of Refs. 10 and 15, as a consequence of the periodicity of the rf field

$$\hat{\mathbf{G}}_i(t+T_{rf},t'+T_{rf})=\hat{\mathbf{G}}_i(t,t'), \tag{22}$$

it is possible to transform expression (16) to the following form:

$$\sigma_i(\omega)=\frac{\sigma_\alpha}{\Gamma_0 T_{rf}}\int_0^{T_{rf}}dt\int_0^{T_{rf}}d\tau\langle W_i(t)| \times\sum_\eta \mathrm{Tr}\left\{\hat V_\eta\frac{\exp(i\tilde\omega\tau)}{\hat{\mathbf{1}}-\exp(i\tilde\omega T_{rf})\hat{\mathbf{G}}_i(t,t+T_{rf})} \times\hat{\mathbf{G}}_i(t,t+\tau)\hat V_\eta^+\right\}|1\rangle+\text{c.c.} \tag{23}$$

This form, where the integration is over finite limits, is more suitable for specific calculations. The total absorption cross section is obviously found by averaging over all particles i with different orientations.

The above result (23) for small amplitudes of the external rf field [condition (15)] cannot be automatically carried over to the case where the external magnetic field exceeds its critical value for some group of particles. Indeed, for magnetic fields exceeding h_c the particle does not have a barrier separating local minima. At these times it is found in one absolute minimum, and the question of relaxation processes at these times becomes moot. In other words, as the external magnetic field varies in time there takes place, as it were, a reduction of states of the particle, between which relaxation occurs. Result (23) can be generalized to this case as follows.

We assume formally, as was done in Sec. 2, that for $|h(t)|>h_c$ the particle is found in two states with identical hyperfine interaction parameters

$$\hat{\mathbf{L}}(t)=\hat{\mathbf{L}}_{\hat H}(t)\otimes\hat 1_e, \tag{24}$$

where $\hat 1_e$ is the unit operator in the space of energy states. Obviously, at the corresponding times relaxation is absent, and in the time interval (t,t') in which the external magnetic field exceeds the critical field the general formula (17) can be used to calculate the function $\hat{\mathbf{G}}_i(t,t')$, where the relaxation operator $\hat P(t)$ can be set equal to zero and the hyperfine interaction operator is given by formula (24).

Let us now consider a time interval (t,t') such that $h(t)<h_c$, $h(t')>h_c$, and there is only one point in time, t_c, at which $h(t_c)=h_c$. According to the SW model, in the point t_c the system undergoes an instantaneous rearrangement. The relaxation process in the vicinity of this point is not described by the relaxation constants, but is prescribed as the final result: in whatever state the particle was found at

times before t_c, at times after t_c it should be found in the state (1). This assumption also remains in force in our model. This process is not hard to take into account if we introduce the projection operator

$$\hat{R}_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}. \tag{25}$$

The operator $\hat{\mathbf{G}}_i(t,t')$ in the indicated time interval obviously has the form

$$\hat{\mathbf{G}}_i(t,t') = \hat{\mathbf{G}}_i(t,t_c)\hat{R}_1\hat{\mathbf{G}}_i(t_c,t). \tag{26}$$

For motion opposite the rf field, when we have $h(t)>h_c$, $h(t')<h_c$, the abrupt rearrangement in the vicinity of the point t_c no longer takes place, and

$$\hat{\mathbf{G}}_i(t,t') = \hat{\mathbf{G}}_i(t,t_c)\hat{\mathbf{G}}_i(t_c,t). \tag{27}$$

In the region of negative values of the rf field, when we have $h(t)>-h_c$, $h(t')<-h_c$, we obtain an expression for the operator $\hat{\mathbf{G}}_i(t,t')$ analogous to expression (26). In this expression the operator $\hat{R}_1$ must be replaced by the projection operator

$$\hat{R}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}. \tag{28}$$

Using these formulas, it is not hard to write out expressions for the superoperator $\hat{\mathbf{G}}_i(t,t')$ even in those cases when the rf magnetic field cycles through the region of critical values repeatedly during the time interval (t,t').

Applying formulas (25)–(28), we can use the general expression (23) to calculate the spectra for arbitrary values of the amplitude of the rf field. Note that the number of calculations grows radically in comparison with the situations considered previously, where either the field varied only in magnitude and not in direction[14] or relaxation was not taken into account.[10]

4. CALCULATIONAL METHODS

The general expression (23) for the absorption spectrum is a double integral in time of complicated superoperator functions, and therefore the numerical realization of the corresponding calculations requires a special analysis both from the point of view of observing the required calculational accuracy and from the point of view of finding the optimal calculation schemes. The central element of these calculations is to calculate the matrix $\hat{\mathbf{G}}(t,t')$ (here and below we drop the index i running over the SW particles). According to the definition of the $\hat{T}$-product,

$$\hat{\mathbf{G}}(t,t') = \lim_{n\to\infty}[\hat{\mathbf{G}}(t,t+\tau)\ldots\hat{\mathbf{G}}(t_k,t_k+\tau)\ldots\hat{\mathbf{G}}(t'-\tau,t')], \tag{29}$$

where $\tau=(t-t')/n$. This definition also gives a basis for calculating the $\hat{\mathbf{G}}(t,t')$ matrices. The calculation reduces to multiplying the n matrices $\hat{\mathbf{G}}(t_k,t_k+\tau)$, which for small τ can be represented in the form of a series in this small parameter:

$$\begin{aligned}\hat{\mathbf{G}}(t_k,t_k+\tau) &= \hat{T}\exp\left\{\int_{t_k}^{t_k+\tau} dt[-\hat{\mathbf{F}}(t)]\right\} \\ &\equiv \hat{\mathbf{1}} - \int_{t_k}^{t_k+\tau} dt_1\hat{\mathbf{F}}(t_1) \\ &\quad + \int_{t_k}^{t_k+\tau} dt_1\int_{t_1}^{t_k+\tau} dt_2\hat{\mathbf{F}}(t_1)\hat{\mathbf{F}}(t_2) - \ldots,\end{aligned} \tag{30}$$

where

$$\hat{\mathbf{F}}(t) = i\hat{\mathbf{L}}(t) + \hat{\mathbf{P}}(t). \tag{31}$$

Through terms in $O(\tau^3)$ we have

$$\hat{\mathbf{G}}(t_k,t_k+\tau) = \hat{\mathbf{1}} - \hat{\mathbf{F}}(t_1)\tau + \frac{1}{2}\hat{\mathbf{F}}(t_1)\hat{\mathbf{F}}(t_1)\tau^2, \tag{32}$$

where $t_1=t_k+\tau/2$. Here the accuracy of calculation of the initial expression (29) will be proportional to τ^2. The necessary calculational accuracy is achieved by increasing the number of partitions n, i.e., by decreasing the corresponding values of τ.

It is clear that the optimal number of partitions n needed to achieve the given calculational accuracy will depend on the ratio of the characteristic frequencies of the problem, which in our case are the frequency of the rf field, the Larmor frequency ω_L in the field H_{hf}, and the relaxation frequencies $p(t)=p_{12}(t)+p_{21}(t)$. Only the direction, not the magnitude of the hyperfine field at the nucleus changes, so that to obtain estimates of the accuracy it can be assumed that the parameter ω_L/ω_{rf} does not vary in time. As for the second parameter, the ratio $p(t)/\omega_{rf}$ can vary over very wide limits. Taking the specifics of the problem into account, specifically that a distinct hyperfine structure should be observed in the absence of the rf field, we assume in what follows that

$$p(h=0) \equiv p_0\exp(-U_0/k_BT) \ll \omega_L/2\pi. \tag{33}$$

Of course, satisfaction of inequality (33) is ensured chiefly by the large value of the barrier U_0, whereas the ratio p_0/ω_L can be arbitrary. Since the dependence $p(t)$ is determined by the size of the barrier and has an exponential character, the inequality

$$p(t) \leqslant \omega_L/2\pi \tag{34}$$

is satisfied over a wide range of times, with the exception of small regions near the critical points. In the interval of times where inequality (34) is satisfied, to calculate the matrix $\hat{\mathbf{G}}(t,t')$ it is necessary to use the above-described procedure with the number of partitions determined by the ratio ω_L/ω_{rf}. On time intervals where condition (34) is not fulfilled, it is necessary to introduce additional partitions.

4.1. Segments with fast relaxation

For large values of the ratio $p(t)/\omega_L$ we expand the superoperator of the hyperfine interaction in Eq. (30) and take the relaxation parameters to be arbitrary:

$$\hat{\mathbf{G}}(t,t+\tau)=\hat{G}_0(t,t+\tau)\oplus\hat{\mathbf{1}}_n + \hat{\mathbf{G}}_1(t,t+\tau)+\hat{\mathbf{G}}_2(t,t+\tau)+\ldots, \tag{35}$$

where

$$\hat{G}_0(t,t+\tau)=\hat{T}\exp\left[-\int_t^{t+\tau}dt'\,\hat{P}(t')\right], \tag{36}$$

$$\hat{\mathbf{G}}_1(t,t+\tau)=\int_0^{\tau}d\tau'\,\hat{\mathbf{G}}_0(t,t+\tau') \times[i\hat{\mathbf{L}}(t+\tau')]\hat{\mathbf{G}}_0(t+\tau',t+\tau), \tag{37}$$

$$\hat{\mathbf{G}}_2(t,t+\tau)=-\int_0^{\tau}d\tau'\int_0^{\tau'}d\tau''\hat{\mathbf{G}}_0(t,t+\tau') \times\hat{\mathbf{L}}(t+\tau')\hat{\mathbf{G}}_0(t+\tau',t+\tau'+\tau'') \times\hat{\mathbf{L}}(t+\tau'+\tau'')\hat{\mathbf{G}}_0(t+\tau'+\tau'',t+\tau). \tag{38}$$

Formulas (35)–(38) define the operator $\hat{\mathbf{G}}(t,t+\tau)$ through terms in $(\omega_L\tau)^3$, where the relaxation parameters do not impose any restrictions.

The relaxation matrix $\hat{P}(t)$ can always be represented in the form

$$\hat{P}(t)=p(t)\hat{S}(t), \tag{39}$$

where

$$\hat{S}(t)=\begin{pmatrix} w_2^{(0)}(t) & -w_2^{(0)}(t) \\ -w_1^{(0)}(t) & w_1^{(0)}(t)\end{pmatrix} \tag{40}$$

is a matrix comprised of the equilibrium populations $w_{1,2}^{(0)}(t)$. If the relaxation parameters vary weakly over times on the order of τ, then

$$\hat{G}_0(t,t+\tau)=\exp[-\hat{P}(t_1)] =\hat{R}(t_1)+\hat{S}(t_1)\exp[-p(t_1)\tau], \tag{41}$$

where $t_1=t+\tau/2$, and

$$\hat{R}(t)=\hat{1}_e-\hat{S}(t)=\begin{pmatrix} w_1^{(0)}(t) & w_2^{(0)}(t) \\ w_1^{(0)}(t) & w_2^{(0)}(t)\end{pmatrix}. \tag{42}$$

With the help of this expression, it is easy to take the integrals on the right-hand sides of Eqs. (37) and (38); as a result we obtain

$$\hat{\mathbf{G}}_1(t,t+\tau)=i\hat{\mathbf{L}}^{RR}\tau+\frac{1-\exp(-p\tau)}{p}i[\hat{\mathbf{L}}^{RS}+\hat{\mathbf{L}}^{SR}] +\tau\exp(-p\tau)i\hat{\mathbf{L}}^{SS}, \tag{43}$$

$$\hat{\mathbf{G}}_2(t,t+\tau)=-(\hat{\mathbf{L}}^{RR})^2\frac{\tau^2}{2}-(\hat{\mathbf{L}}^{RR}\hat{\mathbf{L}}^{RS}+\hat{\mathbf{L}}^{RS}\hat{\mathbf{L}}^{SR} +\hat{\mathbf{L}}^{SR}\hat{\mathbf{L}}^{RR})\frac{\tau}{p}\left[1-\frac{1-\exp(-p\tau)}{p\tau}\right] +(\hat{\mathbf{L}}^{RS}\hat{\mathbf{L}}^{SS}+\hat{\mathbf{L}}^{SR}\hat{\mathbf{L}}^{RS}+\hat{\mathbf{L}}^{SS}\hat{\mathbf{L}}^{SR})\frac{1}{p^2} \times[1-\exp(-p\tau)(1+p\tau)]-(\hat{\mathbf{L}}^{SS})^2 \times\exp(-p\tau)\frac{\tau^2}{2}, \tag{44}$$

where $p\equiv p(t_1)$, $\hat{\mathbf{L}}^{XY}=\hat{X}\hat{\mathbf{L}}\hat{Y}$, and all operators are evaluated at the point t_1. Formulas (35), (41)–(44), on the one hand, do not complicate the calculation excessively and, on the other, make it possible to calculate the matrices $\hat{\mathbf{G}}(t,t')$ on time intervals of a trajectory with fast relaxation without resorting to an extraordinary increase in the number of partitions n. For small $p(t)$ these formulas reduce to the original formula (32).

In the limit $p(t)\tau\gg 1$ through terms in $(\omega_L/\tau)^3$ we have

$$\hat{\mathbf{G}}(t,t+\tau)=\hat{R}(t_1)\otimes\exp(i\overleftarrow{\mathbf{L}}_n(t_1)\tau), \tag{45}$$

where

$$\overleftarrow{\mathbf{L}}_n(t)=w_1^{(0)}(t)\hat{\mathbf{L}}_{\hat{H}}^{(1)}(t)+w_2^{(0)}(t)\hat{\mathbf{L}}_{\hat{H}}^{(2)}(t) \tag{46}$$

is the mean superoperator of the hyperfine interaction averaged over equilibrium states.

4.2. Integration over time with additional factorization

Let us turn to the main formula (23) and estimate the number of operations needed to calculate the absorption spectrum. As a simple estimate we take the total number N_{tot} of products of two complex numbers, neglecting the number of sums, which require substantially less calculation time. Simple estimates show that

$$N_{\text{tot}}\propto N_iN_\omega n^2N^2, \tag{47}$$

where N_i is the number of partitions into groups with different orientations, N_ω is the number of points at which the Mössbauer spectrum is calculated, n is the number of partitions needed to calculate the matrices $\hat{\mathbf{G}}_i(t,t')$, and N is given by formula (21). For a reasonable choice of values, namely, $N_i\approx 128$, $N_\omega\approx 512$, $n\approx 256$, $N=16$, we obtain the number of required complex multiplications $N_{\text{tot}}\approx 2^{40}\approx 10^{12}$. Obviously, to perform these calculations on a computer in this case requires a large amount of time. In reality, the calculation time can be substantially simplified.

Returning to the original formula (16) for the absorption cross section, we divide the integral on its right-hand side into two parts:

$$\sigma_i(\omega)=\sigma_i^{(1)}(\omega)+\sigma_i^{(2)}(\omega), \tag{48}$$

where

$$\sigma_i^{(1)}(\omega)=\frac{\sigma_\alpha}{\Gamma_0T_{rf}}\int_0^{T_{rf}}dt\int_{T_{rf}}^{\infty}dt'\langle W_i(t)| \times\sum_\eta\,\mathrm{Tr}\{\hat{V}_\eta\hat{\mathbf{G}}_i(t,t') \times\exp[i\tilde{\omega}(t'-t)]\hat{V}_\eta^+\}|1\rangle+\text{c.c.}, \tag{49}$$

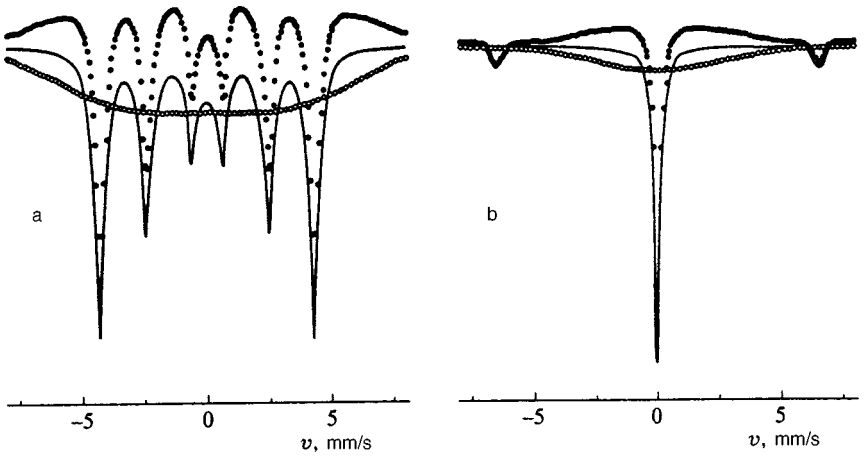

FIG. 6. Mössbauer spectra (solid lines) and partial components of the absorption cross section $\sigma_i^{(1)}(\omega)$ and $\sigma_i^{(2)}(\omega)$ (empty and filled points, respectively) for an ensemble of SW particles ($U_0/k_BT=20$, $p_0/(\omega_{rf}/2\pi)=10^4$) in an rf field with $\omega_{rf}/2\pi=75$ MHz, $h_0=0.2$ (a) and 0.3 (b). The scale of spectrum b is reduced by a factor of four.

$$\sigma_i^{(2)}(\omega)=\frac{\sigma_\alpha}{\Gamma_0 T_{rf}}\int_0^{T_{rf}}dt\int_t^{T_{rf}}dt'\langle W_i(t)|\times\sum_\eta \mathrm{Tr}\{\hat V_\eta\hat{\mathbf{G}}_i(t,t')\times\exp[i\tilde\omega(t'-t)]\hat V_\eta^+\}|1\rangle+\text{c.c.} \tag{50}$$

From definition (17) of the matrix $\hat{\mathbf{G}}(t,t')$ as a $\hat T$-product it follows that it possesses the property

$$\hat{\mathbf{G}}(t,t')=\hat{\mathbf{G}}(t,T_{rf})\hat{\mathbf{G}}(T_{rf},t') \quad \text{for} \quad t'>T_{rf}>t. \tag{51}$$

Employing relation (51), we can reduce expression (49) to the form

$$\sigma_i^{(1)}(\omega)=\frac{\sigma_\alpha}{\Gamma_0 T_{rf}}\int_0^{T_{rf}}dt\int_0^{T_{rf}}dt'\langle W_i(t)|\times\sum_\eta \mathrm{Tr}\left\{\hat V_\eta(t)\frac{\exp[i\tilde\omega(t'-t)]\exp(i\tilde\omega T_{rf})}{\hat{\mathbf{1}}-\exp(i\tilde\omega T_{rf})\hat{\mathbf{G}}_i(0,T_{rf})}\times\hat V_\eta^+(t')\right\}|1\rangle+\text{c.c.}, \tag{52}$$

where

$$\hat V_\eta(t)=\hat V_\eta\hat{\mathbf{G}}_i(t,T_{rf}),\quad \hat V_\eta^+(t)=\hat{\mathbf{G}}_i(0,t)\hat V_\eta^+. \tag{53}$$

In turn, making the substitution of variables $\tau=t'-t$ also reduces the integral on the right-hand side of Eq. (50) to a form more convenient for calculation:

$$\sigma_i^{(2)}(\omega)=\frac{\sigma_\alpha}{\Gamma_0}\int_0^{T_{rf}}d\tau\exp(i\tilde\omega\tau)I_i(\tau)+\text{c.c.}, \tag{54}$$

where

$$I_i(\tau)=\frac{1}{T_{rf}}\int_0^{T_{rf}-\tau}dt\langle W_i(t)|\times\sum_\eta \mathrm{Tr}\{\hat V_\eta\hat{\mathbf{G}}_i(t,t+\tau)\hat V_\eta^+\}|1\rangle. \tag{55}$$

It can be seen by inspection that expression (52) factorizes so that the double integral in t and t' reduces to the calculation of two single integrals. Here, of course, the number of operations needed to calculate expression (52) is sharply decreased and instead of expression (47), taking the averaging over groups of particles with different orientations into account, we have the estimate

$$N_{\text{tot}}\propto N_i N_\omega n N^2. \tag{56}$$

For the same values of N_i, N_ω, n, and N we obtain for the number of required complex multiplications $N_{\text{tot}}\approx 2^{32}\approx 10^9$, which provides a basis for carrying out the calculations using moderate computing resources, specifically personal computers.

As for the second term $\sigma_i^{(2)}(\omega)$, its calculation requires a considerably smaller number of calculations than does the calculation of $\sigma_i^{(1)}(\omega)$. Indeed, the main fraction of the calculations are needed in the calculation of the correlation function $I(\tau)$. It is not hard to show that $N_i n^2 N^2$ complex multiplications suffice for calculations taking into account averaging over groups of particles with different orientations. Since the same calculations do not need to be repeated for different points of the spectrum, the given estimate is small in comparison with estimate (56).

For high frequencies of the rf field, when we have $\omega_{rf}>\omega_L$, the main contribution to the absorption cross section comes from the term $\sigma_i^{(1)}(\omega)$. The contribution of the second term is a smooth curve with characteristic variations at frequencies near ω_{rf} (see Fig. 6). In the limit of high frequencies ω_{rf} it is possible to obtain the following expression for this term:

$$\sigma_i^{(2)}(\omega)=\frac{2\sigma_\alpha T_{rf}}{\Gamma_0}\frac{1-\cos(\omega T_{rf})}{(\omega T_{rf})^2}. \tag{57}$$

However, it should be noted that even if the contribution from $\sigma_i^{(2)}(\omega)$ is small, it cannot be neglected, since the contribution from $\sigma_i^{(1)}(\omega)$ can give a physically invalid result with negative values of the absorption cross section exceeding the background (see Fig. 6). The second term compensates for these negative contributions, so that the resulting absorption cross section is always positive.

In the region of low frequencies of the rf field, when we have $\omega_{rf}\ll\omega_L$, the main contribution to the absorption cross section comes from $\sigma_i^{(2)}(\omega)$, and the first term gives only small corrections.

Figure 5 (right side) displays relaxation Mössbauer absorption spectra for different amplitudes of the rf field, calculated according to the above-described scheme. As in the case of the original SW model, the well-resolved hyperfine structure collapses into a central line with satellites as the amplitude of the rf field is increased. Here, as in Fig. 5, collapse occurs at considerably smaller amplitudes of the rf field in comparison with the original model, i.e., relaxation results in an effective decrease in the magnitude of the critical field (see Sec. 2 and Fig. 4). Also, the spectral lines on which a hyperfine structure is still observed are found to be broadened and their width grows with the amplitude of the field. It is clear that for a quantitative description of the spectra it is very important to have a more accurate description of the relaxation processes. As for qualitative effects, if they can show up they can be expected to appear only in the region of the transition from a resolved hyperfine structure to a central line with satellites (see Figs. 5b–5d). It may be expected that the abrupt qualitative change in the shape of the spectra for values of the amplitude of the rf field near the lower value of the critical fields $h_c=0.5$ will be smoothed out by the relaxation processes. An analysis of the transformation of Mössbauer spectra in this initial stage of the transition as a function of the problem parameters is of special interest.

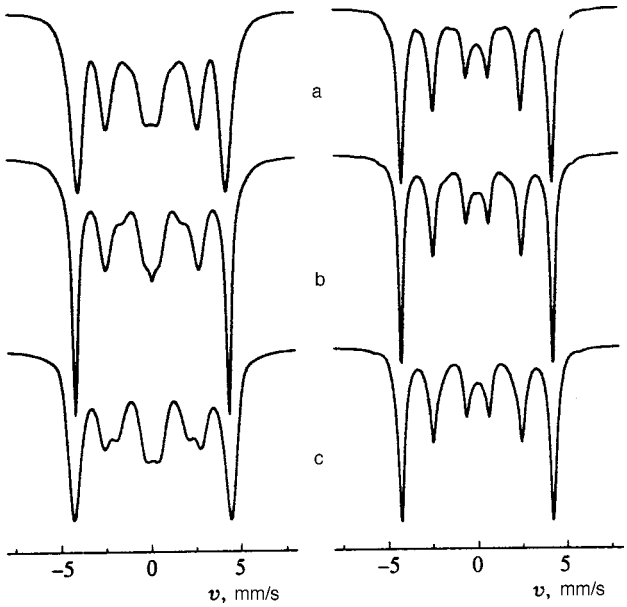

FIG. 7. Mössbauer spectra of SW particles with orientation $\theta=45°$ (left) and an ensemble of SW particles (right) for $U_0/k_BT=20$, $p_0/(\omega_L/2\pi)=10^4$ in an rf field with amplitude $h_0=0.2$ and frequency near the frequency of the hyperfine component $\omega_1/2\pi=50$ MHz ($\Delta\omega/2\pi=5$ MHz): $\omega_{rf}=\omega_1+\Delta\omega$ (a), ω_1 (b), $\omega_1-\Delta\omega$ (c). The scale of the spectra on the right has been reduced by a factor of 1.8.

5. RELAXATION-STIMULATED RESONANCES

The proposed relaxation model differs from the original SW model in one very important physical way. In the original model the remagnetization process is understood in two ways: 1) if the amplitude of the rf field h_0 exceeds the critical field h_c, then remagnetization occurs every half-period of the rf field, and 2) for $h_0<h_c$ remagnetization does not occur and the particle returns to its initial state every half-period. The generalized relaxation model admits such remagnetization regimes, where during a half-period the particle can return to to its initial state with probability r or change the direction of its magnetic moment with probability $q=1-r$. Qualitative effects in the Mössbauer spectra can be expected precisely in these remagnetization regimes.

Calculations of the spectra as functions of the frequency of the rf field and the relaxation parameters in the above-described scheme showed that in the case of slow relaxation ($q\ll 1$) the behavior of the spectra has a distinctive character that depends on the relative values of the frequencies of the hyperfine structure and the frequency of the rf field. Since the minimum value of the critical field $h_c=0.5$ is realized for the group of particles with $\theta=45°$, to reveal the specifics of the transition we first analyzed the absorption spectra for this group of particles. Figure 7 shows the corresponding spectra for representative values of the frequency of the rf field near

$$\omega_{rf}=\omega_1, \tag{58}$$

where $\omega_1=|\pm 3\omega_e/2\mp\omega_g/2|$ is the frequency corresponding to lines 1 and 6 in the magnetic sextet ($\omega_{e,g}=g_{g,e}\mu_N H_{rf}$). For exact resonance, as can be seen from the figure, the outer lines are much narrower than the inner ones and, consequently, their peak intensity is dramatically increased. As one goes away from the resonance toward either side, the line width is restored, where the inner part of the spectrum varies relatively weakly. This is effect is so strong that it is preserved even in spectra for the entire ensemble of SW particles.

A qualitatively different situation arises for frequencies of the rf field in the vicinity of

$$\omega_{rf}=2\omega_1. \tag{59}$$

The corresponding spectra are shown in Fig. 8. For exact resonance, splitting of the outer lines of the sextet is observed, which can be interpreted as a superposition of the central line, e.g., line 1, and a satellite of the second line of the pair, line 6. This interpretation finds confirmation in the shape of the spectra calculated for the case of small detuning of the frequency of the rf field from the ''resonance'' frequency (Figs. 8a and 8c). In these spectra one can clearly make out the satellites of lines 1 and 6, indicated by arrows in the figure. Note that in this case no satellites of the inner lines of the spectrum are observed, and that satellites of lines 1 and 6 are absent in the inner regions of the spectrum. Equally interesting is the fact that lines 1 and 6 and the corresponding satellites can never coincide for any values of the frequency of the rf field, so there exists some minimal distance Δ_{12} between the lines. As a result, we cannot say which lines in Fig. 8b are the central lines, and which are satellites.

Interest in resonance phenomena in the presence of an rf field in Mössbauer spectroscopy arose in the 1960's and has not weakened to the present day.[12,13] However, attempts to find such effects were concentrated on cases of resonance with real distances between the levels of the nucleus in the ground state (ω_g) or the excited state (ω_e). The latter should

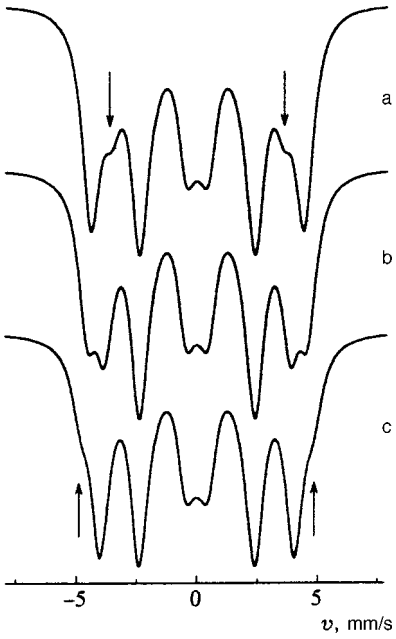

FIG. 8. Transformation of the Mössbauer spectra of SW particles with orientation $\theta=45°$ for $U_0/k_BT=20$, $p_0/(\omega_L/2\pi)=10^4$ in an rf field with amplitude $h_0=0.2$ and frequency near twice the frequency of the hyperfine component $\omega_1/2\pi=50$ MHz: $\omega_{rf}/2\pi=104$ (a), 99 (b), 94 MHz (c).

appear in the form of splitting of all components of the spectrum, and they show up distinctly in the SW model [see Fig. 9a, and also Ref. 10]. As can been seen from Fig. 9b, a more accurate account of the relaxation process causes a strong smearing of this effect, which may explain the fact that the numerous efforts to detect these resonance effects have long been unsuccessful.

Thus, relaxation processes hinder one from observing ordinary physical resonances, but stimulate the appearance of new resonances now at frequencies of the hyperfine transitions instead of the nuclear transitions. The nontrivial character of these resonances arises because the frequencies of the hyperfine transitions exceed the actual distances between the nuclear levels by ten orders of magnitude.

Obviously, it is practically impossible to track the physics of the formation of these resonance phenomena on the basis of general formulas (52)–(55), which not only do not have such a simple analytical form, but also require significant efforts just for their computer realization. Nevertheless, we may suggest a simplified model that would make it possible to some degree to explain this phenomenon.

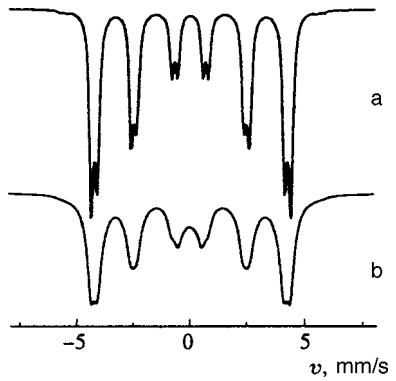

FIG. 9. Mössbauer spectra of an ensemble of SW particles with amplitude $h_0=0.2$ at the resonant frequency $\omega_{rf}/2\pi=\omega_g/2\pi=36.8$ MHz in the original SW model (a) and in the extended relaxation model for $U_0/k_BT=20$ and $p_0/(\omega_L/2\pi)=10^4$ (b).

For simplicity, we assume that the amplitude of the rf field is not large, so that variations of the direction of the hyperfine field at the nucleus in states corresponding to local energy minima can be neglected. In this case, we can transform from expressions (52)–(55) with complicated operator functions to a sum over the various frequencies of the hyperfine transitions (see Ref. 14); as a result, we obtain

$$\sigma(\omega)=\sigma_\alpha\sum_\alpha |C_\alpha|^2\varphi_\alpha(\omega), \tag{60}$$

where $\alpha=(m,M)$; m and M are the projections of the nuclear spin in the ground state and the excited state onto the direction of the hyperfine field; the coefficients C_α determine the intensities of the corresponding hyperfine transitions,

$$\varphi_\alpha(\omega)=\frac{1}{\Gamma_0}\langle A_\alpha(\omega)|\frac{T_{rf}}{\hat 1-\exp(i\tilde\omega T_{rf})\hat g_\alpha(0,T_{rf})}|B_\alpha(\omega)\rangle + \int_0^{T_{rf}} d\tau\exp(i\tilde\omega\tau)I_\alpha(\tau)+\text{c.c.}, \tag{61}$$

$$\langle A_\alpha(\omega)|=\frac{1}{T_{rf}}\int_0^{T_{rf}} dt\langle W(t)|\hat g_\alpha(t,T_{rf})\exp[i\tilde\omega(T_{rf}-t)], \tag{62}$$

$$|B_\alpha(\omega)\rangle=\frac{1}{T_{rf}}\int_0^{T_{rf}} dt\exp(i\tilde\omega t)\hat g_\alpha(0,t)|1\rangle, \tag{63}$$

$$I_\alpha(\tau)=\frac{1}{T_{rf}}\int_0^{T_{rf}-\tau} dt\langle W(t)|\hat g_\alpha(t,t+\tau)|1\rangle. \tag{64}$$

Here $\hat g_\alpha(t,t')$ is an operator that is represented by a matrix of second rank:

$$\hat g_\alpha(t,t')=\hat T\exp\left[\int_t^{t'} dt''(-\hat\omega_\alpha-\hat P(t''))\right], \tag{65}$$

$$\hat\omega_\alpha=\begin{pmatrix}\omega_\alpha & 0\\ 0 & -\omega_\alpha\end{pmatrix}, \tag{66}$$

where $\omega_\alpha=M\omega_e-m\omega_g$.

A quirk of the above relaxation process is the fact that for an overwhelming number of particles, with the exception of the group of particles oriented perpendicular to the direction of the rf field, relaxation has a unilateral character, specifically, probabilities of transitions from a higher energy state to a lower state are of considerable significance while the probabilities of the reverse transitions are negligibly small. Moreover, even according to formulas (6) and (7) the relaxation process is concentrated at the ends of the trajectories $t_c=kT_{rf}/2$, where the intensity of the rf field reaches its maximum value, and the relaxation rate falls off with distance from these points. (Here, for simplicity, we reckon time from the point where the intensity of the rf field has its minimum value, i.e., $h(0)=-h_0$.) We assume that the re-

laxation process is maximally localized, but that there is a finite integral effect, defined by the constants r and q:

$$r=\exp\left[-2\int_{T_{rf}/2-t_\varepsilon}^{T_{rf}/2}p(t)dt\right],\quad q=1-r, \tag{67}$$

where t_ε defines the time interval where intense relaxation processes occur; according to our assumption $t_\varepsilon\to 0$.

In the vicinity of these points, in the calculation of the functions $\hat{g}_\alpha(t,t')$ it is possible to neglect the hyperfine interactions and keep only the relaxation operator, in which case we have

$$\hat{g}_\alpha(T_{rf}/2-t_\varepsilon,T_{rf}/2+t_\varepsilon)=\hat{1}-\begin{pmatrix}0&0\\-q&q\end{pmatrix},$$

$$\hat{g}_\alpha(T_{rf}-t_\varepsilon,T_{rf}+t_\varepsilon)=\hat{1}-\begin{pmatrix}q&-q\\0&0\end{pmatrix}. \tag{68}$$

For intermediate times, when t and t' lie in the intervals $(t_\varepsilon,T_{rf}/2-t_\varepsilon)$, $(T_{rf}/2+t_\varepsilon,T_{rf}-t_\varepsilon)$, relaxation processes can be neglected, and taking properties (27) into account we easily find for the function $\hat{g}_\alpha(0,T_{rf})$

$$\hat{g}_\alpha(0,T_{rf})=\begin{pmatrix}r\exp(-i\omega_\alpha T_{rf})&q\exp(-i\omega_\alpha T_{rf})\\qr&q^2+r\exp(i\omega_\alpha T_{rf})\end{pmatrix}. \tag{69}$$

We assume that the relaxation is weak, i.e.,

$$q\ll 1. \tag{70}$$

We will consider segments of the spectrum near the frequency ω_α, so that

$$|\omega-\omega_\alpha|\ll\omega_\alpha. \tag{71}$$

We also assume that the frequency of the rf field is near one of the parametric-resonance frequencies:

$$|\omega_{rf}-\omega_n^{(r)}|\ll\omega_\alpha, \tag{72}$$

where

$$\omega_n^{(r)}=2\omega_\alpha/n. \tag{73}$$

If conditions (70) and (71) are satisfied, the vectors (62) and (63) do not depend on the frequency and take on a simple form:

$$\langle A_\alpha(\omega)|=\frac{1}{2}(1,0),\quad |B_\alpha(\omega)\rangle=\begin{pmatrix}1\\0\end{pmatrix}. \tag{74}$$

As for the second term in formula (61), it gives only a small background contribution (see Sec. 4.2). It is not hard to write out an expression for this term in explicit form, but within the limits of accuracy of formula (74) we should set this term identically equal to zero, since an account of terms linear in q in the calculation of vectors (74) would give corrections comparable with the contribution of the second term. Moreover, in the given approximation we should also discard terms quadratic in q in the term $\hat{g}_\alpha(0,T_{rf})$ in formula (61).

By utilizing formulas (74) and (69) while keeping only terms linear in q, we can reduce formula (61) to the simple form

$$\varphi_\alpha(\omega)=-\frac{1}{\Gamma_0}\times\mathrm{Im}\frac{\omega-\omega_\alpha-\Delta\omega+i\Gamma/2}{(\omega-\omega_\alpha+i\Gamma/2)(\omega-\omega_\alpha-\Delta\omega+i\Gamma/2)\pm\gamma^2}, \tag{75}$$

where

$$\gamma=q/T_{rf}, \tag{76}$$

$$\Delta\omega=n\omega_{rf}-2\omega_\alpha, \tag{77}$$

$$\Gamma=\Gamma_0+2\gamma. \tag{78}$$

The + and − signs before the second term in the denominator in expression (75) correspond to odd and even resonances (73). It is clear that expression (75) can be represented as a sum of two lines with Lorentzian shape:

$$\varphi_\alpha(\omega)=-\frac{1}{\Gamma_0}\mathrm{Im}\left(\frac{A_1}{\omega-\omega_\alpha-\lambda+i\Gamma/2}+\frac{A_2}{\omega+\omega_\alpha-n\omega_{rf}+\lambda+i\Gamma/2}\right), \tag{79}$$

where

$$\lambda=\frac{1}{2}(\sqrt{(\Delta\omega)^2\mp 4\gamma^2}-|\Delta\omega|), \tag{80}$$

$$A_1=\frac{|\Delta\omega|+\lambda}{|\Delta\omega|+2\lambda},\quad A_2=\frac{\lambda}{|\Delta\omega|+2\lambda}. \tag{81}$$

If the relaxation process is not important and it is possible to set $\gamma=0$, we have one Lorentzian line at the frequency ω_α. But as soon as the relaxation is switched on, a second line appears, located at the frequency $-\omega_\alpha+n\omega_{rf}$, i.e., at the position of a satellite of the hyperfine component with frequency $-\omega_\alpha$. Thus, the relaxation process generates the appearance of satellites. Moreover, as follows from formulas (79)–(81), such generation has a sharp resonance character. At large detunings of ω_{rf} from the resonant frequency, when $|\Delta\omega|\gg 2\gamma$, the intensity of the central line is close to unity and the intensity of the satellite is low. On the other hand, at exact resonance, as can be seen from formula (81), the intensities of both lines are equal: $A_1=A_2=0.5$.

It is interesting to track the transformation of the shape of these lines. In the given case, we have two different types of behavior depending on the parity of the resonance. Thus, for even resonances [see formula (58)] at exact resonances the quantity λ turns out to be purely imaginary, so that both lines coincide in position, but differ in width:

$$\Gamma_1=\Gamma_0,\quad \Gamma_2=\Gamma_0+4\gamma. \tag{82}$$

For $\gamma\gg\Gamma_0$ the spectrum is a superposition of a narrow line and a wide line, which should be manifested as an abruptincrease in the peak intensity for these lines, as we demonstrated by numerical calculations based on the general formulas (52)–(55) (see Fig. 7). As one moves away from the exact resonance, the width of the first line increases while the

width of the second line decreases, for $|\Delta\omega|=2\gamma$ they are equal, and with further growth of $|\Delta\omega|$ the line widths do not change.

Qualitatively different behavior obtains in the case of odd resonances [see formula (59)]. In this case, the quantity λ is always real and the widths of the two lines are identical. But on the other hand, as follows from formulas (79) and (80), these lines can never coincide, i.e., between them there is a minimum distance

$$\Delta_{12}=2\gamma, \tag{83}$$

and they cannot lie closer than that (see Fig. 8).

Thus, the above analysis within the framework of the simplified relaxation model allows one to reveal in clearer form aspects of the numerically detected relaxation-stimulated resonance effects. Note that the analysis within the framework of the simplified model cannot replace exact calculations, which take into account the presence of particles with different orientations and variations of the hyperfine field in direction. Thus, our calculations based on the exact formulas for the entire ensemble of SW particles show that even resonances show up quite clearly not only in the Mössbauer spectra for a certain group of particles with a given orientation, but also in the spectra of the entire ensemble, which gives hope of the possibility of their eventual experimental detection. On the other hand, for odd resonances all of these properties are preserved for individual groups of particles, whereas for the ensemble of particles they are, to a significant degree, ''washed out'' and for this reason do not show up in Fig. 8. Clearly, the experimental detection of odd resonances will require the preparation of textured samples.

6. CONCLUSION

In the present paper we have constructed a theory of Mössbauer absorption spectra in the presence of a rf magnetic field for a system of SW particles in the extended relaxation model, which allows a more complete description of the relaxation process. In this model it is possible to describe hysteresis loops, which depend on the frequency of the rf field, and thereby explain the huge differences between the remagnetization fields observed in ordinary magnetic measurements and obtained from the Mössbauer absorption spectra. Qualitatively new behavior in the absorption spectra is predicted in the region of rf fields where a resolved hyperfine structure is still observed, consisting in relaxation-stimulated processes of intense satellite lines at frequencies of the rf field coupled by parametric-resonance conditions with the hyperfine frequencies. These effects should be observed in weak rf fields, and the conditions for their observation should be realized in experiments in a simpler way than the conditions for the observation of the collapse effect.

APPENDIX A

TABLE I. Matrix representation of the superoperator $\hat{\mathbf{L}}_{\hat{H}}(t)$ for ^{57}Fe nuclei.

$\lvert M\rangle\langle m\rvert$	$\lvert\frac{3}{2}\rangle\langle\frac{1}{2}\rvert$	$\lvert\frac{3}{2}\rangle\langle-\frac{1}{2}\rvert$	$\lvert\frac{1}{2}\rangle\langle\frac{1}{2}\rvert$	$\lvert\frac{1}{2}\rangle\langle-\frac{1}{2}\rvert$	$\lvert-\frac{1}{2}\rangle\langle\frac{1}{2}\rvert$	$\lvert-\frac{1}{2}\rangle\langle-\frac{1}{2}\rvert$	$\lvert-\frac{3}{2}\rangle\langle\frac{1}{2}\rvert$	$\lvert-\frac{3}{2}\rangle\langle-\frac{1}{2}\rvert$
$\lvert\frac{3}{2}\rangle\langle\frac{1}{2}\rvert$	$\frac{3\omega_e-\omega_g}{2}\times\cos\phi$	$-\frac{\omega_g}{2}\sin\phi$	$\frac{\sqrt{3}\omega_e}{2}\sin\phi$	0	0	0	0	0
$\lvert\frac{3}{2}\rangle\langle-\frac{1}{2}\rvert$	$-\frac{\omega_g}{2}\sin\phi$	$\frac{3\omega_e+\omega_g}{2}\times\cos\phi$	0	$\frac{\sqrt{3}\omega_e}{2}\sin\phi$	0	0	0	0
$\lvert\frac{1}{2}\rangle\langle\frac{1}{2}\rvert$	$\frac{\sqrt{3}\omega_e}{2}\sin\phi$	0	$\frac{\omega_e-\omega_g}{2}\times\cos\phi$	$-\frac{\omega_g}{2}\sin\phi$	$\omega_e\sin\phi$	0	0	0
$\lvert\frac{1}{2}\rangle\langle-\frac{1}{2}\rvert$	0	$\frac{\sqrt{3}\omega_e}{2}\sin\phi$	$-\frac{\omega_g}{2}\sin\phi$	$\frac{\omega_e+\omega_g}{2}\times\cos\phi$	0	$\omega_e\sin\phi$	0	0
$\lvert-\frac{1}{2}\rangle\langle\frac{1}{2}\rvert$	0	0	$\omega_e\sin\phi$	0	$-\frac{\omega_e+\omega_g}{2}\times\cos\phi$	$-\frac{\omega_g}{2}\sin\phi$	$\frac{\sqrt{3}\omega_e}{2}\sin\phi$	0
$\lvert-\frac{1}{2}\rangle\langle-\frac{1}{2}\rvert$	0	0	0	$\omega_e\sin\phi$	$-\frac{\omega_g}{2}\sin\phi$	$-\frac{\omega_e-\omega_g}{2}\times\cos\phi$	0	$\frac{\sqrt{3}\omega_e}{2}\sin\phi$
$\lvert-\frac{3}{2}\rangle\langle\frac{1}{2}\rvert$	0	0	0	0	$\frac{\sqrt{3}\omega_e}{2}\sin\phi$	0	$-\frac{3\omega_e+\omega_g}{2}\times\cos\phi$	$-\frac{\omega_g}{2}\sin\phi$
$\lvert-\frac{3}{2}\rangle\langle-\frac{1}{2}\rvert$	0	0	0	0	0	$\frac{\sqrt{3}\omega_e}{2}\sin\phi$	$-\frac{\omega_g}{2}\sin\phi$	$-\frac{3\omega_e-\omega_g}{2}\times\cos\phi$

Remark. Here $\phi\equiv\phi(t)$.

[1] L. Pfeiffer, J. Appl. Phys. **42**, 1725 (1971).
[2] L. Pfeiffer, N. D. Heiman, and J. C. Walker, Phys. Rev. B **6**, 74 (1972).
[3] G. Asti, G. Albanese, and C. Bucci, Phys. Rev. **184**, 260 (1969).
[4] M. Kopcewicz, J. Phys. (Paris), Colloq. **37**, C6-107 (1976).
[5] M. Kopcewicz, U. Gonser, and H.-G. Wagner, Nucl. Instrum. Meth. **199**, 163 (1982).
[6] M. Kopcewicz, Hyperfine Interact. **71**, 1453 (1992).
[7] T. Graf, M. Kopcewicz, and J. Hesse, Nanostruct. Mater. **6**, 937 (1995).
[8] T. Graf, M. Kopcewicz, and J. Hesse, J. Phys.: Condens. Matter **8**, 3897 (1996).
[9] J. Hesse, T. Graf, M. Kopcewicz, *et al.*, Hyperfine Interact. **113**, 499 (1998).
[10] A. M. Afanas'ev, M. A. Chuev, and Yu. Khesse, Zh. Éksp. Teor. Fiz. **113**, 1799 (1998) [JETP **86**, 983 (1998)].
[11] E. C. Stoner and E. P. Wohlfarth, Philos. Trans. R. Soc. London, Ser. A **240**, 599 (1948).
[12] E. Matthias, in *Hyperfine Interactions and Nuclear Transitions*, edited by E. Matthias and D. A. Shirley (North-Holland, Amsterdam, 1968), p. 815.
[13] F. G. Vagizov, R. A. Manapov, E. K. Sadykov, and L. L. Zakirov, Hyperfine Interact. **116**, 91 (1998).
[14] A. M. Afanas'ev, M. A. Chuev, and J. Hesse, Phys. Rev. B **56**, 5489 (1997).
[15] W. Heitler, *Quantum Theory of Radiation* (Clarendon Press, Oxford, 1954).
[16] R. Zwanzig, Physica (Utrecht) **30**, 1109 (1964).
[17] P. W. Anderson, J. Phys. Soc. Jpn. **9**, 316 (1954).

Translated by Paul F. Schippnick

Calculation of the field dependence of the rates of emission of carriers from deep centers based on an experimental form-function for the optical transition

S. V. Bulyarskiĭ,*) N. S. Grushko, and A. V. Zhukov

Ulyanovsk State University, 432700 Ulyanovsk, Russia
(Submitted 28 December 1998)
Zh. Éksp. Teor. Fiz. **116**, 1027–1034 (September 1999)

An algorithm is proposed for calculating the field dependence of the emission rates based on a form-function for the optical transition. A comparison is made with experimental data for the $V_{Ga}S_{As}$ complex in gallium arsenide. This scheme for calculating the field dependence is found to be preferable to methods based on a single-coordinate model. © *1999 American Institute of Physics.* [S1063-7761(99)01809-0]

Multiphoton processes play an important role in radiationless transitions.[1–4] Electron–phonon interactions give rise to temperature dependence for capture coefficients and to an increase in the thermal emission rate in strong electric fields. In particular, it has been shown experimentally[5] and theoretically[6–14] that in strong electric fields, the probability of these transitions increases exponentially as the square of the electric field strength. Thus, more accurate calculations of the parameters of optoelectronic and high power semiconductor devices associated with generation, recombination, and tunnelling processes involving deep levels will require information on the parameters of the electron–phonon interactions which characterize a given electronic transition.

The theory has found practical application mainly for the single-coordinate model.[6–14] This model imposes rather rigid requirements on the character of the oscillations in the system and requires verification in each individual case. Degeneracy of the electronic states of a crystal with an impurity center causes the adiabatic approximation to fail and gives rise to vibrational mixing of the electronic levels. In this case, the single-coordinate model may not be suitable for calculating the field dependence.

In this paper we propose an algorithm for calculating the field dependence of the thermal emission rates which relies on a form-function for the optical transition calculated from the emission spectrum. Some results obtained in this way are compared with experiment.

Gallium arsenide doped with sulfur was chosen as the material for testing the model experimentally. Group VI impurities in gallium arsenide occupy arsenic sites and become donors, forming shallow levels near the bottom of the conduction band. In addition, they are known[15–17] to form ''gallium-donor vacancy complexes at arsenic sites'' ($V_{Ga}D_{As}$). These complexes produce broad bands with peaks at 1.18–1.25 eV in the photoluminescence spectrum. The symmetry and electronic structure of these complexes have been studied in detail[15–17] and it is also known that in an excited state a hole captured by one of these complexes interacts with incompletely symmetric vibrations of the atoms surrounding V_{Ga}, i.e., the Jahn–Teller effect occurs.

The samples of GaAs:S were grown by gaseous phase epitaxy with concentrations of dopant impurity ranging from $4.7\cdot10^{17}\,\mathrm{cm}^{-3}$ to $1.5\cdot10^{18}\,\mathrm{cm}^{-3}$. The photoluminescence spectral characteristics were recorded over wavelengths from 800 to 1200 nm and temperatures from 100 to 200 K on an SDL-2M system.

The spectra from all the samples had two characteristic peaks: one with photon energies at the peak of $h\nu_{\max}=1.48-1.49\,\mathrm{eV}$ and the second, a broader peak with $h\nu_{\max}=1.22-1.24\,\mathrm{eV}$. The first luminescence peak was associated with band–band emission in GaAs and the second, with emission from $V_{Ga}S_{As}$ complexes [Fig. 1(a)]. Given that ''gallium-donor vacancies in the nearest lattice site'' produce deep levels in the gap of GaAs near the ceiling of the valence band, the luminescence band with a peak at 1.22–1.24 eV corresponds to radiative recombination of an electron in a state near the bottom of the conduction band with a hole localized at the deep center being studied.

A study of the dependence of the integrated emission intensity of these complexes on the dopant impurity concentration showed that the intensity actually is proportional to the concentration of sulfur in the samples, which confirms the model for the complex that has been chosen [Fig. 1(b)]. It was also found that the shape of the emission bands (in particular the dispersion of the band) at fixed temperatures is independent of the dopant impurity concentration and is probably determined by an electron–phonon interaction, rather than by doping effects. Measurements at different temperatures [Fig. 1(a)] confirm this. As the temperature is raised, a shift to lower energies and a broadening of the emission band are observed. The shape of the spectra and their temperature dependence suggest that the electron–phonon interaction plays an important role.

Metal–semiconductor contacts were fabricated on the test samples by electrochemical deposition of nickel on the gallium arsenide. A study of the electrical properties of these contacts showed that the charge carrier transport is described by thermionic emission. Thus, the contacts served as Schottky barriers with a potential barrier height of 1 eV. We studied the field dependence of the rate of thermal emission of holes from a deep level created by the $V_{Ga}S_{As}$ complex in these structures. The experiments were done at 77 K in the

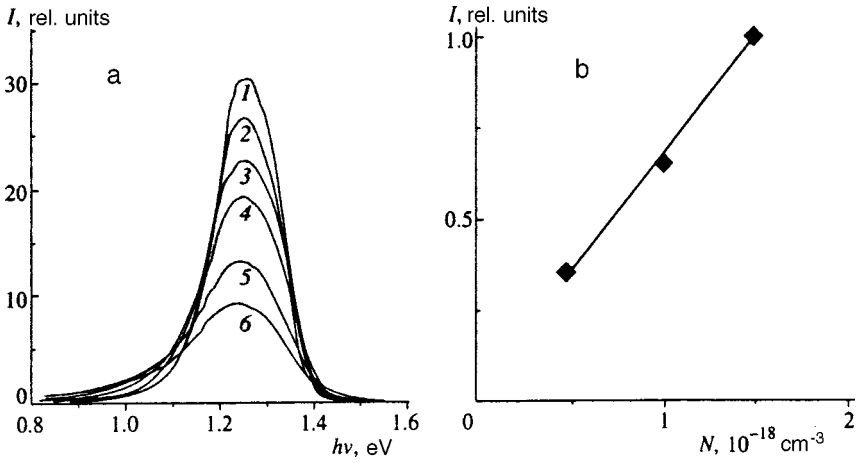

FIG. 1. a) Emission spectra of $V_{Ga}S_{As}$ complexes at different temperatures: *1*—100 K; *2*—118 K; *3*—137 K; *4*—161 K; *5*—180 K; *6*—200 K. b) Integrated intensity of the emission spectrum of $V_{Ga}S_{As}$ complexes as a function of dopant impurity concentration.

following way. The centers were photoionized using AL-106 infrared photodiodes with a peak emission at a wavelength of $\lambda = 914$ nm ($h\nu = 1.36$ eV, which corresponds roughly to the peak in the absorption spectrum of $V_{Ga}S_{As}$). The photocapacitance kinetics was measured when the light was turned on and turned off.

The analysis of the measurement data relied on a simple kinetic equation, which takes the following form in the case where there is no capture of electrons and holes in the field of the space-charge region:

$$\frac{dn_t}{dt} = -(Jq_n + e_n)n_t + (Jq_p + e_p)(N_t - n_t), \tag{1}$$

where J is the flux of photons in the space-charge region, $q_{n(p)}$ is the electron (hole) cross section for photoionization, $e_{n(p)}$ is the rate of emission of electrons (holes) from the level, N_t is the concentration of the complexes, and n_t is the concentration of electrons in the complexes. The emission rate is a combination of all thermal-field processes. Thus, the time constant for the fall in capacitance when the light is turned off is $\tau^{-1} = e_n + e_p$.

Given that the level lies closer to the valence band and the energy distance to the bands is more than $10kT$, we may assume that the recharging of the level when the light is turned off is entirely determined by hole emission and that the time constant for this process is given by $\tau^{-1} = e_p$. The experiment was repeated for different reverse bias voltages, so it was possible to find the field dependence of the thermal emission rate. Typical plots of the emission rate as a function of the square of the field in the space-charge region are shown in Fig. 2. We note, once again, that in this case the emission rate we have measured includes a combination of all the thermal-field processes and is essentially the probability of a hole transition from a deep level.

The calculations of the field dependence relied on Ref. 18, in which it was shown rigorously that the probability of a quantum mechanical transition including the electron–phonon interaction can be written, in general, in terms of the overlap integral (see the Appendix):

$$W = \sum_{i,j} \int_{-\infty}^{\infty} W_{i,j}^{(0)}(E_{i,j}^{(t)} - \varepsilon) f_{i,j}(\varepsilon) d\varepsilon, \tag{2}$$

where $W_{i,j}^{(0)}(E_{i,j}^{(t)} - \varepsilon)$ is the probability of a purely electronic transition, $E_{i,j}^{(t)}$ is the energy of the purely electronic transition from sublevel i of the multiplet of the initial state of the center to sublevel j of the final state of the multiplet, and $f_{i,j}(\varepsilon)$ is the form-function for this optical transition. Essentially, $f_{i,j}(\varepsilon)$ represents the contribution of the electron–phonon transition to the transition probability.

The probability of a purely electronic transition can be calculated using the formula:[19]

$$W_{i,j}^{(0)}(E_{i,j}^{(t)} - \varepsilon) = A \frac{eF}{2\sqrt{2m^*(E_{i,j}^{(t)} - \varepsilon)}} \times \exp\left(-\frac{4}{3} \frac{\sqrt{2m^*(E_{i,j}^{(t)} - \varepsilon)^3}}{e\hbar F} \right), \tag{3}$$

where A is a normalizing factor, F is the electric field strength, and m^* is the effective electron mass.

In order to calculate the field dependence of the transition probability using Eq. (2), it is necessary to know the form-function for absorption by an electron from a local state near the valence band to a level. We shall find this form-

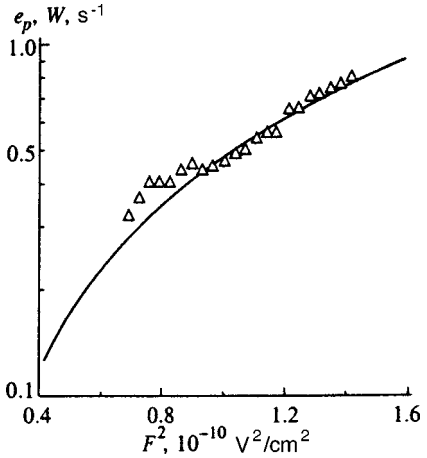

FIG. 2. Experimental (e_p, triangles) and theoretical (W, smooth curve) field dependence of the probability for a transition of a hole from deep levels corresponding to the ground state of a $V_{Ga}S_{As}$ center to a local state near the ceiling of the valence band.

function in the following way: we calculate the form-function for the radiative transition and then its moments. We then transform them, relying on the model for the complex and the symmetry of the wave functions.

If the emission spectrum of an electronic transition is known, then its form-function can be calculated using the formula[20]

$$f(\varepsilon)=I(\varepsilon)/M_0\varepsilon^4. \tag{4}$$

Here $I(\varepsilon)$ is the luminescence intensity, ε is the photon energy, and M_0 is the zeroth moment of the emission band, which is proportional to the oscillator strength, and is given by $M_0=\int I(\varepsilon)d\varepsilon$.

We now consider the sum associated with the degeneracy of the electronic terms in Eq. (2).

According to the model proposed for $V_{Ga}Te_{As}$ complexes in Refs. 15–17, because of the Jahn–Teller effect there is a reduction in the symmetry of the complex in its excited state where each defect continues to exist in one of three equivalent configurations corresponding to the three possible orientations of the Jahn–Teller distortion. In each of these configurations, there exists a {110} symmetry plane which contains the original axis of the complex, i.e., the lattice sites corresponding to the initial position of both components of the complex. A reorientation of the Jahn–Teller distortions of the complex reduces to a rotation of this plane about the initial axis by an angle $\varphi=\pm 2\pi/3$. It has been shown rigorously[21] that in this case the adiabatic potentials of the three terms of the excited state are energetically and geometrically equivalent with respect to the ground-state equilibrium configuration of the complex. Thus, the probabilities of the purely electronic transitions are equal for each of the configurations. In this case, we can take this probability out from under the summation sign in Eq. (2), placing the sum immediately in front of the form-functions:

$$W=\int_{-\infty}^{\infty} W^{(0)}(E^{(t)}-\varepsilon)\sum_{i,j} f_{i,j}(\varepsilon)d\varepsilon. \tag{5}$$

Then $\Sigma_{i,j}f_{i,j}(\varepsilon)=f(\varepsilon)$ and this quantity is calculated from experimental data using Eq. (4).

Equation (5), therefore, contains a form-function derived from experimental emission spectra (the form-function of a transition from a local state near the bottom of the conduction band corresponding to an excited state of the $V_{Ga}S_{As}$ complex into the ground state) which makes it possible to perform the numerical integration and determine the transition probability.

The form-function for a transition from a local state near the ceiling of the valence band to a deep level corresponding to the ground state of $V_{Ga}S_{As}$ (for which the field dependence of the thermal emission rate have been measured) is calculated on the basis of the following arguments.

Strictly speaking, models of the electronic–vibrational interaction are valid for intracenter transitions.[22] Thus, we shall assume that the wave functions are determined only by the states of the recombination center.

According to Ref. 16, the center we are studying has a nondegenerate ground state, whose wave function has s-symmetry (we shall denote the wave function of this state by $|s\rangle$) and a degenerate excited state formed of p-type wave functions. We shall consider two transitions: one that is radiative from a local state near the bottom of the conduction band, corresponding to an excited state of the center, into the ground state and another with absorption from a local state near the ceiling of the valence band to a deep level corresponding to the ground state of the center. Given that there is a fairly strong electron–phonon interaction in the system (the experimentally determined magnitude of the Stokes loss is ~ 0.12 eV) and that both initial states (near the bottom of the conduction band and the ceiling of the valence band) correspond to the same excited state of the complex, we can assume that the wave functions of these states are the same, i.e., $|p\rangle=|p'\rangle$. The energy spectrum of this system consists of one s-level, which corresponds to the ground state of the center with energy E_s, and two groups of close levels p_i and p_i' with energies E_{pi} and $E_{p'i}$.

The form-functions for the bands with emission from state $|p\rangle$ into state $|s\rangle$ and absorption from state $|p'\rangle$ into state $|s\rangle$, respectively, can be written in the form[22,23]

$$f^{(e)}(\varepsilon)=\frac{1}{z_p}\sum_i \sum_{s,p} |\langle s|\hat{M}|p_i\rangle|^2 \exp\left(-\frac{E_{pi}}{kT}\right)\delta(E_s-E_{pi}+\varepsilon), \tag{6}$$

and

$$f^{(a)}(\varepsilon)=\frac{1}{z_{p'}}\sum_i \sum_{s,p'} |\langle s|\hat{M}|p_i'\rangle|^2 \times\exp\left(-\frac{E_{p'i}}{kT}\right)\delta(E_s-E_{p'i}-\varepsilon), \tag{7}$$

where $\Sigma_{s,p(p')}$ is a sum over the vibrational states of the s and $p(p')$ terms,

$$z_{p(p')}=\sum_i \exp\left(-\frac{E_{p(p')}}{kT}\right)$$

is the partition function of the group of levels $p(p')$, and $\hat{M}$ is the perturbation operator that generates the transition.

Given that $|p\rangle=|p'\rangle$ and $E_{p'i}=E_{pi}-E_g$, where E_g is the band gap, we rewrite Eq. (7) in the form

$$f^{(a)}(\varepsilon)=\frac{1}{z_p}\sum_i \sum_{s,p} |\langle s|\hat{M}|p_i\rangle|^2 \times\exp\left(-\frac{E_{pi}}{kT}\right)\delta(E_s-E_{pi}-\varepsilon+E_g). \tag{8}$$

Equation (8) is the same as Eq. (6) after transformation to the new variable $\varepsilon'=-\varepsilon+E_g$ in Eq. (8).

Therefore, by taking the mirror reflection of our experimental emission form-function with respect to the $\varepsilon=0$ axis and shifting to higher energies by an amount E_g, we obtain the form-function of the absorption band for a transition from a local state near the ceiling of the valence band to a deep level corresponding to the ground state of the center. We substitute this form-function in Eq. (2), taking Eq. (5) into account, and calculate the probability of hole emission from a deep level of the $V_{Ga}S_{As}$ complex as a function of the average field in the space-charge region. This dependence is

plotted in Fig. 2 (smooth curve). The factor A was chosen for a best fit of the calculated and experimental curves at high fields.

Therefore, a calculated field dependence has been obtained without making the assumptions associated with the single-coordinate model. The basis for these calculations is the experimental form-function, which naturally reflects the complex electron–vibrational interactions. In this regard, a computational scheme of this type for calculating the field dependence is preferable to methods based on the single-coordinate model.

APPENDIX

Here we provide a brief derivation of Eq. (2). The transition probability under the influence of a perturbation $\hat{H}'$ is given by[23]

$$W=\sum_{n,n'}\rho_{1n}|\langle 1_n|\hat{H}'|2_{n'}\rangle|^2\delta(E_{2n'}-E_{1n}-E^{(t)}), \tag{9}$$

where n and n' enumerate the vibrational states of the ground and excited electronic terms and ρ_{1n} is the probability of finding an electron in the vibrational state n of term 1,[23] which, for a Boltzmann distribution, is given by

$$\rho_{1n}=\frac{\exp(-E_{1n}/kT)}{\Sigma_{n''}\exp(-E_{1n''}/kT)}.$$

In the adiabatic approximation the quantum mechanical state vector can be written in the form $|2_{n'}\rangle=|2^e\rangle|2^L_{n'}\rangle$, where $|2^e\rangle$ is the state vector of the electron and $|2^L_{n'}\rangle$ is the state vector of the lattice. In first-order perturbation theory[24] the vector $|2^e\rangle$ is independent of the coordinate Q of the oscillator (Condon approximation). Thus, we can partition the matrix element into a purely electronic component and a purely vibrational component which contains only the overlap integral of the oscillator wave functions:

$$W=\sum_{n,n'}\rho_{1n}|\langle 1^e|\hat{H}'|2^e\rangle|^2|\langle 1^L_n|2^L_{n'}\rangle|^2\delta(E_{2n'}-E_{1n}-E^{(t)}). \tag{10}$$

We shall use the rule

$$\int_{-\infty}^{\infty}F(x)\,\delta(y-x)\,dx=F(y).$$

We then obtain

$$\begin{aligned}W(E^{(t)})&=\int_{-\infty}^{\infty}\sum_{n,n'}\rho_{1n}|\langle 1^e|\hat{H}'|2^e\rangle|^2|\langle 1^L_n|2^L_{n'}\rangle|^2\delta\\&\quad\times(E_{2n'}-E_{1n}-\varepsilon)\,\delta(\varepsilon-E^{(t)})\,d\varepsilon\\&=\int_{-\infty}^{\infty}|\langle 1^e|\hat{H}'|2^e\rangle|^2\delta(\varepsilon-E^{(t)})\\&\quad\times\sum_{n,n'}\rho_{1n}|\langle 1^L_n|2^L_{n'}\rangle|^2\delta(E_{2n'}-E_{1n}-\varepsilon)\,\delta(\varepsilon)\\&=\int_{-\infty}^{\infty}W^{(0)}(E^{(t)}-\varepsilon)g(\varepsilon)\,d\varepsilon,\end{aligned} \tag{11}$$

where $W^{(0)}(E^{(t)}-\varepsilon)$ is the probability of a purely electronic transition and $g(\varepsilon)$ is a function containing information on the contribution of the phonon subsystem to the transition probability.

We shall show that, to within a constant coefficient, $g(\varepsilon)$ equals the form-function $f(\nu)$ of the optical transition, which depends on the frequency ν of the emitted light:

$$\begin{aligned}f(\nu)&=\sum_{n,n'}\rho_{1n}|\langle 1_n|\hat{M}|2_{n'}\rangle|^2\delta(E_{2n'}-E_{1n}-h\nu)\\&=\sum_{n,n'}\rho_{1n}|\langle 1^e|\hat{M}|2^e\rangle|^2|\langle 1^L_n|2^L_{n'}\rangle|^2\delta(E_{2n'}-E_{1n}-h\nu)\\&=|\langle 1^e|\hat{M}|2^e\rangle|^2\sum_{n,n'}\rho_{1n}|\langle 1^L_n|2^L_{n'}\rangle|^2\delta(E_{2n'}-E_{1n}-h\nu)\\&=|M^0_{12}(h\nu)|^2g(h\nu).\end{aligned} \tag{12}$$

Now Eq. (11) can be rewritten in the form

$$W(E^{(t)})=\int_{-\infty}^{\infty}W^{(0)}(E^{(t)}-\varepsilon)\frac{f(\varepsilon)}{|M^0_{12}(\varepsilon)|^2}\,d\varepsilon. \tag{13}$$

Within the limits of the optical band of the $1\rightarrow 2$ transition, the matrix element of the dipole–dipole interaction can be regarded as independent of energy,[21] so

$$W(E^{(t)})=\int_{-\infty}^{\infty}W^{(0)}(E^{(t)}-\varepsilon)f(\varepsilon)\,d\varepsilon. \tag{14}$$

Here the matrix element of the dipole interaction appears in the normalization coefficient in $f(\varepsilon)$.

We now consider the case where the energy spectrum consists of two groups of close levels $(1_i;2_j)$, separated by a large energy gap. Suppose the relaxation time within group 1 is considerably shorter than the lifetime with respect to $1\rightarrow 2$ transitions. Then[22] the expression for the transition probability can be written in the form

$$W=\sum_{i,j}\int_{-\infty}^{\infty}W^{(0)}_{i,j}(E^{(t)}_{i,j}-\varepsilon)f_{i,j}(\varepsilon)\,d\varepsilon, \tag{15}$$

where the sum is taken over all sublevels of the multiplets 1 and 2.

*)E-mail: har@ulsu.ru

[1] Huang Kun and A. Rees, in *Problems in the Physics of Semiconductors* [Russian translation], Gostekhizdat, Moscow (1957), 389 pp.
[2] S. I. Pekar, Usp. Fiz. Nauk **50**, 197 (1953).
[3] V. A. Kovarskiĭ, *Kinetics of Radiationless Processes* [in Russian], Izd-vo. Kishinevskogo Un-ta, Kishinev (1968).
[4] H. Kukimoto and C. H. Henry, Phys. Rev. B **7**, 2486 (1973).
[5] S. V. Bulyarskiĭ and N. S. Grushko, Fiz. Tekh. Poluprovodn. **9**, 287 (1975) [Sov. Phys. Semicond. **9**, 187 (1975)].
[6] D. Pons and S. Makram-Ebeid, J. de Phys. **40**, 1168 (1979).
[7] S. Makram-Ebeid and M. Lannoo, Phys. Rev. **25**, 6406 (1982).
[8] S. F. Timashov, Fiz. Tverd. Tela **14**, 2621 (1972) [Sov. Phys. Solid State **14**, 2267 (1972)].
[9] S. F. Timashov, Fiz. Tverd. Tela **14**, 171 (1972) [Sov. Phys. Solid State **14**, 136 (1972)].
[10] V. Karpus and V. I. Perel', JETP Lett. **42**, 497 (1985).
[11] V. Karpus, JETP Lett. **44**, 430 (1986).

[12] V. Karpus and V. I. Perel', Zh. Éksp. Teor. Fiz. **91**, 2319 (1986) [Sov. Phys. JETP **64**, 1376 (1986)].
[13] V. N. Abakumov, I. A. Merkulov, V. I. Perel', and I. N. Yassievich, Zh. Éksp. Teor. Fiz. **89**, 1472 (1985) [Sov. Phys. JETP **62**, 853 (1985)].
[14] V. N. Abakumov, O. V. Kurnosova, A. A. Pakhomov, and I. N. Yassievich, Fiz. Tverd. Tela **30**, 1793 (1988) [Sov. Phys. Solid State **30**, 1030 (1988)].
[15] N. S. Averkiev, A. A. Gutkin, E. B. Osipov *et al.*, Fiz. Tekh. Poluprovodn. **25**, 50 (1992) [Sov. Phys. Semicond. **25**, 28 (1992)].
[16] N. S. Averkiev, A. A. Gutkin, E. B. Osipov *et al.*, Fiz. Tekh. Poluprovodn. **25**, 57 (1992) [Sov. Phys. Semicond. **25**, 33 (1992)].
[17] A. A. Gutkin, M. A. Reshchikov, and V. E. Sedov Fiz. Tekh. Poluprovodn. **31**, 1062 (1997) [Semiconductors **31**, 908 (1997)].
[18] Yu. B. Rozenfel'd, S. V. Bulyarskiu, and E. P. Evseev, in *Abstracts of Talks at the XIII-th All-union Conf. on Semiconductor Theory* [in Russian], Izd-vo. Erevanskogo Un-ta, Erevan (1987), 281 pp.
[19] W. Franz, in S. Flügge (Ed.), *Handbuch der Physik*, Vol. 17 (Springer–Verlag, Berlin, 1956) [Russian transl. *Breakdown in Dielectrics*, IL, Moscow (1961)].
[20] K. K. Rebane, A. P. Purga, and O. I. Sil'd, Trudy Instituta Fiziki i Astronomii AN ÉSSR, No. 14, p. 31 (1961).
[21] Yu. E. Perlin, Usp. Fiz. Nauk **80**, 553 (1964) [Sov. Phys. Usp. **6**, 542 (1964)].
[22] Yu. E. Perlin and B. S. Tsukerblat, *Effects of Electron-Vibrational Interactions in Optical Spectra of Impurity Paramagnetic Ions* [in Russian], Stiintsa, Kishinev (1974).
[23] I. B. Berskuer, *Electronic Structure and Properties of Transition Metal Compounds*, Wiley–Interscience, New York (1996) [Russian orig., Khimiya, Leningrad (1976), 350 pp.].
[24] B. Ridley, *Quantum Processes in Semiconductors*, 2nd ed., Clarendon Press, Oxford (1988) [Russian transl. Mir, Moscow (1986)].

Translated by D. H. McNeill

Transient photoimpedance response of $YBa_2Cu_3O_{7-\delta}$ epitaxial films in the mixed state

A. D. Semenov*)

Moscow State Pedagogical University, 119435 Moscow, Russia
(Submitted 29 January 1999)
Zh. Éksp. Teor. Fiz. **116**, 1035–1047 (September 1999)

This paper reports on an investigation of transient photoimpedance response to radiation, in other words, real-time variations of the impedance induced by femtosecond optical pulses in superconducting films transferred to the mixed state by an external magnetic field applied parallel to the *c*-axis. When the films were in a state characterized by the absence of *dc* resistivity, the response amplitude increased with the magnetic field faster than expected owing to the contribution of magnetic vortices to the impedance of a superconductor with *s*-wave pairing of electrons. It turned out that the effect is due to a growth in the effective density of quasiparticle states in the mixed state of a *d*-wave superconductor. In the absence of magnetic field, however, the response amplitude was higher at lower temperatures, which contradicts the predictions of both models. Possible reasons for this feature in the response as a function of temperature are discussed. © *1999 American Institute of Physics.* [S1063-7761(99)01909-5]

1. INTRODUCTION

Several different mechanisms of the photoimpedance response of superconducting films based on interaction between optical photons and magnetic vortices have recently been put forward. These include hops of vortices between neighboring pinning centers driven by electromagnetic radiation (flux creep),[1] viscous flux flow,[2] and generation of vortex-antivortex pairs with their subsequent dissociation.[3] As far as we know, none of these mechanisms has yet been detected in experiments. One complication for experimental studies of such effects is the lack of a model that would describe interaction of an optical photon with a magnetic vortex as a whole without perturbing the surrounding region of the superconductor.

Resonant interaction between photons of lower energies in the terahertz band and quasi-particles localized around vortices has been detected[4] in YBaCuO films. Optical photons, whose energy is considerably higher than both the superconducting gap width and the characteristic binding energy of quasiparticles in vortex cores, likewise interact with Cooper pairs and quasiparticles. In this case, the description of the nonthermal impedance response of a film in the superconducting state in terms of real-time variations in the superconducting condensate density is quite satisfactory,[5] and in the resistive state it can be accounted for in terms of the effective electron temperature and temperature dependence of the film resistivity around the superconducting transition point.[6] A photoresponse due to optically driven relaxation of the flux line density gradient has been detected in YBaCuO films in Ref. 7. The authors of the latter study, however, assert that the primary effect of optical radiation was film heating. Even so, the presence of vortices can indirectly affect the photoresponse, since the film impedance in the superconducting state changes in the presence of magnetic vortices.

This paper reports on the photoimpedance response of YBaCuO epitaxial films in the mixed state generated by applied magnetic field. The measured response as a function of magnetic field has turned out too steep to be ascribed to the contribution of magnetic vortices in an isotropic *s*-wave superconductor. This dependence is put down to modifications in the quasiparticle density of states in the mixed state of a *d*-wave superconductor.

2. EXPERIMENTAL TECHNIQUES

YBaCuO epitaxial films with a thickness of 90 nm were grown on lanthanum aluminate ($LaAlO_3$) substrates so that their *c*-axis is perpendicular to the substrate plane. From these films, we lithographically fabricated structures shaped as bridges 50 μm long and 5 μm wide connected to gold-coated contact pads. The sample resistivity at the room temperature was $7\times10^{-6}\,\Omega\cdot$m and dropped linearly with decreasing temperature so that it was $3\times10^{-6}\,\Omega\cdot$m at a temperature slightly higher than the superconducting transition temperature. The transition temperature was 92 K, and its width was 1.8 K. At the liquid-nitrogen temperature, the critical current density in the samples was $1.8\times10^{6}\,\mathrm{A\cdot cm^{-2}}$. The samples were placed in an optical cryostat which allowed us to vary the temperature between 4 K and room temperature, and to apply a *dc* magnetic field of up to 4 T. The magnetic field was generated by superconducting Helmholtz coils and aligned with the *c*-axis of the film.

The laser system, which included a pulsed titanium-sapphire laser with active phase locking and a pulse selector, generated a sequence of pulses with a duration of 100 fs at a wavelength of 0.8 μm and a rate of 2.6 MHz. The system for detection of optically induced signals included high-frequency spring-loaded contacts, an electric bias-tee, a broad-band amplifier, and a sampling oscilloscope. The transient time of the entire circuit was 37 ps. The laser beam was

focused on the sample surface into a spot 40 μm in diameter. Accurate measurements of the radiation power absorbed by the sample were performed using the same sample, whose temperature was maintained, with this end in view, in the middle of the superconducting transition, and which acted in this case as a bolometer with parameters (temperature response, response time) measured in advance. The time constant of the bolometric response of our samples was 13 ns, which is in conformity with both the film thickness and our previously published results.[8] This allowed us to determine the radiation energy density absorbed by the sample per pulse, which proved to be 15 $\mu J \cdot cm^{-2}$. In measurements of the photoresponse versus magnetic field, the samples were cooled to the required temperature at zero magnetic field.

When an optical pulse is absorbed by a sample, the impedance of the latter changes. If a bias current flows across the sample, a pulsed electrical signal $S(t)$ is generated between its terminals. The signal $E(t)$ observed on the oscilloscope is related to the amplitude and shape of the voltage transient across the sample; nonetheless, it is not completely determined by the latter, but also depends on the impulse response of the electric circuit. Voltage pulses across the sample with a duration shorter than the circuit transient time are distorted so that neither their shape nor duration are faithfully reproduced by the oscilloscope.

The situation is much simpler if the shape of the pulse across the sample is constant. In this case, the signal amplitude is proportional (with a constant factor) to the maximal deviation of the sample impedance from its equilibrium value. Moreover, if the maximal variation of the impedance is small in comparison with its equilibrium value, and the time-dependent current component is much smaller than the bias, the measured signal amplitude M can be expressed as follows:

$$M = \max[E(t)] = \max\left[\int_0^t S(t)K(t-\tau)d\tau\right] = k\max[S(t)],$$

$$S(t) = I\left[R(t) - \frac{d}{dt}L(t)\right]. \tag{1}$$

Here $K(t)$ is the impulse response of the detection circuit, I is the bias current, and $R(t)$ and $L(t)$ are the time-dependent sample resistance and inductance. The factor k depends on both the electric pulse shape across the sample and the impulse response of the electric circuit. This factor, however, is invariant under the conditions listed above. Since the impedance of a superconducting film in the mixed state is a complex function of frequency, the formulas become more compact if we use the Fourier transform of Eq. (1):

$$S_\omega = I(R_\omega - j\omega L_\omega) = I[\mathrm{Re}(Z_\omega) - j\,\mathrm{Im}(Z_\omega)], \tag{2}$$

where Z_ω is the complex frequency-dependent sample impedance and ω is the angular frequency. In the limiting case of a thin film, $d \ll \lambda_L$, where d is the film thickness and λ_L is the London magnetic field penetration depth, the impedance can be expressed in terms of the complex magnetic field penetration depth λ_ω as $Z_\omega = j\omega\mu_0(\lambda_\omega)^2/d$. In this case, Eq. (2) reduces to

$$S_\omega = \omega\mu_0 I \frac{l}{wd}\{\mathrm{Im}(\lambda_\omega^2) + j\,\mathrm{Re}(\lambda_\omega^2)\}, \tag{3}$$

where μ_0 is the magnetic permeability of the vacuum and l, w, and d are the length, width, and thickness of the sample. In order to analyze experimental results, we calculated on the basis of various theories changes in λ_ω due to absorption of optical pulses, then calculated the signal maxima using the inverse Fourier transform and compared them with measured signal amplitudes.

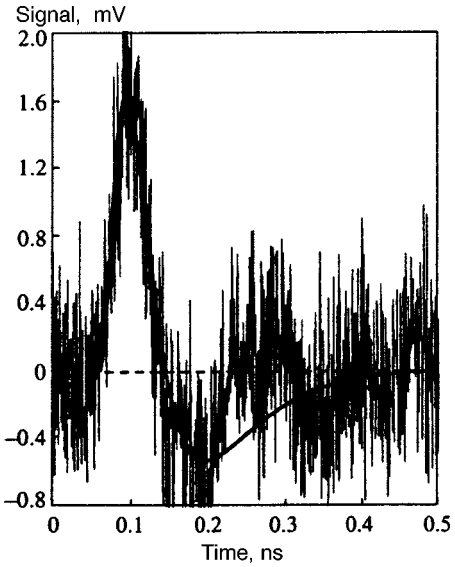

FIG. 1. Typical photoresponse signal recorded at zero magnetic field at a temperature of 70 K and bias current of 1 mA. The solid line shows model calculations. The dashed line is the zero signal level.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Experimental results

At all temperatures below the superconducting transition, we recorded typical bipolar optically induced signals (see Ref. 6 and references therein) if the sample *dc* resistivity was zero, and both the current bias and optical energy density were sufficiently low. Such curves were described in literature many times and interpreted in terms of a change of the kinetic inductance of a superconducting film under irradiation. A typical signal recorded in our experiment is shown in Fig. 1 with a curve obtained using the formalism described in the previous section (the fitting parameters will be discussed later). The signal is composed of a symmetrical positive component of width 75 ps, and a negative component whose decay time is slightly longer than the rise time. The oscillations in the region of the negative component decay are probably due to the impedance mismatch between the sample and high-frequency contacts.

The positive signal amplitude, which is defined as the maximal signal measured with respect to the base line, is proportional to the optical pulse energy and bias current

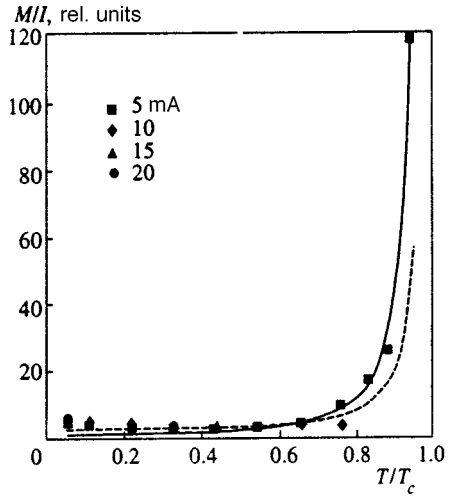

FIG. 2. Response amplitude normalized to the bias current versus temperature. Measurements were performed in zero magnetic field at various currents. The solid and dashed lines plot calculations for the *s*-wave and *d*-wave pairing, respectively.

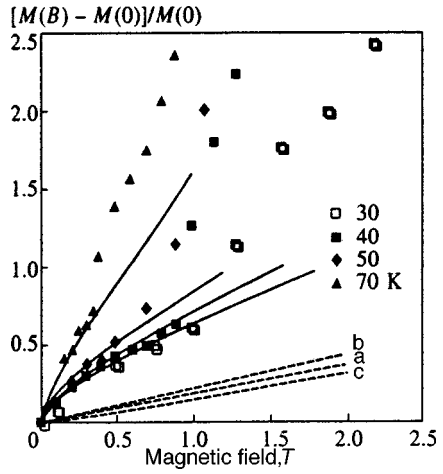

FIG. 3. Normalized response amplitudes at different temperatures as functions of magnetic field. The solid lines plot calculations by Eq. (9) with fitting parameters adjusted to obtain the best agreement between calculations and measurements. The dashed lines plot calculations for an *s*-wave superconductor at a temperature of 50 K. In the latter case, for better visibility, the calculated signal amplitudes were multiplied by a factor of five. The following limiting cases were analyzed: a) vortices are immobile; b) vortices are retained at pinning centers but capable of oscillating; c) in addition to oscillations, vortices can contribute to the creep and flux flow.

when both these parameters are fairly small and the signal shape is constant. In what follows, we will term the regime of signal detection in this region of parameters a linear regime. When the current and/or pulse energy go beyond the boundaries of this region, the signal has a bolometric component whose decay time is considerably longer, of the order of several nanoseconds. The emergence of the purely thermal component is correlated with the onset of *dc* resistivity, although the latter event occurred at somewhat higher currents or energy density in the pulse. After feeding a sufficiently high current (usually several tens of milliamperes) through a sample cooled in zero magnetic field, we observed a weak response in the absence of the bias current with the sign opposite to that in the presence of a bias current. When the bias current was cyclically varied between a negative and a positive value, the signal exhibited hysteresis, and the loop width increased as the temperature was lowered.

In the absence of magnetic field, the positive amplitude of the linear response dropped considerably with decreasing temperature (Fig. 2), but below the temperature of about $0.35T_c$ and down to the lowest temperature attainable in the experiment of 4 K the signal grew slightly as the temperature dropped. An important point is that, although changes in the signal in the range of low temperatures were small, they were notably larger than the experimental uncertainty and reproducible. This property of the response was observed at all bias currents that were within the region of the linear regime. The growth in the response amplitude at low temperatures is a very interesting feature, given that the amplitude of the kinetic response is assumed to be inversely proportional to the superconducting condensate density in the simplest models. Under magnetic field, the signal amplitude was, generally speaking, higher, although hysteresis was detected in the amplitude versus magnetic field (up to about 0.2 T) when the field increased from zero to a certain value and then dropped to zero. The magnetic field dependence of the normalized amplitude of the positive component is shown in Fig. 3.

At each temperature, the bias current was selected sufficiently small so as on the one hand to have a linear response at the maximal magnetic field, and on the other, to have a responsivity sufficient for detecting a signal in zero magnetic field. In the region of intermediate magnetic fields, the signal amplitude was a sublinear function of the field strength, whereas the signal shape was invariant. At higher magnetic fields, there was an inflection point on the curves of the signal versus magnetic field at each temperature. The slope in the region below the inflection point increased with the temperature. The magnetic field magnitude at the inflection point marked the boundary between the linear regime at lower magnetic fields and the regime in which a bolometric component was detected and samples had a nonvanishing *dc* resistivity. Since the bolometric response of HTSC films has been studied previously in detail and is beyond the scope of this paper, here we concentrate on the linear response, which, given such characteristics as its low amplitude and fast decay, can be associated beyond doubt with nonthermal mechanisms. The slope of the normalized signal amplitude as a function of magnetic field (0.6–1.5 per tesla) in the region of intermediate fields is obviously higher than predictions based on theories developed for superconductors with an isotropic order parameter and *s*-wave electron pairing. Indeed, in this case a drop in the superconducting condensate density due to generation of magnetic flux lines and the corresponding increase in the normalized signal amplitude should be of order $B/B_{c2}(T)$, which is 0.05 per tesla even if we take the highest second critical field for YBaCuO cited in literature. In the next section, we will discuss how the derivative of the signal with respect to magnetic field intensity can change if the electron pairing and superconducting gap have different spatial symmetry properties.

3.2. Interaction between optical photons and magnetic vortices

In early publications,[1,2] the possibility of depinning an isolated magnetic vortex through its interaction with an optical photon was intuitively associated with the proximity between two characteristic energies, namely, the photon energy and depinning potential. These two parameters, however, have different physical natures. The photon energy is a microscopic parameter characterizing interaction between the photon and an electron, whereas the depinning potential is a thermodynamic parameter characterizing the binding energy of a vortex as a whole to a pinning center and is used in describing the dynamic equilibrium between free and bound vortices at a finite temperature. Thus, a correct description of magnetic vortex depinning by one photon requires a microscopic mechanism that would redistribute the photon energy among all electrons forming the flux line or, in other words, screening the magnetic flux associated with the vortex. With the exception of the local heating mechanism discussed in a series of publications by Kadin and co-workers (see, for example, Ref. 3), no such mechanism has been suggested in the above cited papers and subsequent publications. In the case of depinning due to local heating, the critical time scale is determined by the time of electron thermalization τ_{th}, which is of order 1 ps in YBaCuO (Ref. 9) is essentially independent of temperature. Here we do not take into account phonon heating, which is significantly slower and leads to a much smaller increase in the temperature.

Since the optical pulse width in our experiment is much smaller than τ_{th}, one can say that all photons of one optical pulse are absorbed simultaneously. The material heating is local when the average distance between photons, $a_\lambda = (h\nu/\varepsilon_0)^{1/2}$, where $h\nu$ is the photon energy and ε_0 is the energy density in the laser pulse, is considerably larger than the electron diffusion length during the thermalization time, $L_T=(D\tau_{th})^{1/2}$, where D is the electron diffusion coefficient. In the opposite limit, $L_T>a_\lambda$, photons absorbed by the sample lead to a uniform heating of electrons. Estimates based on the average temperature derivative of the second critical field at the transition temperature, 0.5 T/K,[10] yield an electron diffusion coefficient of 3 $cm^2\cdot s^{-1}$, and accordingly a diffusion length of 150 Å, which is considerably larger than both the vortex core diameter and the distance between layers of the crystal lattice along the *c*-axis. The latter dimension determines the minimal length of the flux line fragment which can move in a quasi-independent manner. Thus, the local heating can have a reasonable efficiency in generating additional vortices capable of flowing through the sample in the viscous manner only if the average separation between photons equals the average distance between the vortices and the density of the latter is fairly low.

However attractive the concept of optical depinning may seem, the pulse energy density of 0.1 $\mu J/cm^2$, which marks the boundary between the regimes of local and uniform heating, is much lower than the sensitivity achieved in modern experiments with femtosecond laser pulses. For this reason, we assume in discussions of the reported experiments that the electron heating is uniform and that the effective electron temperature and superconducting condensate density are uniquely related to one another by time-independent equation over a time interval from τ_{th} to the time of electron-phonon interaction. The moment when the positive signal component achieves its maximum is within this time interval.

In the absence of magnetic vortices, the magnetic field penetration depth is constant with the frequency over the range limited by the reciprocal laser pulse width and equals the conventional London penetration depth $\lambda_L \approx f^{1/2}$, where f is the relative density of the superconducting condensate. The impedance response to laser pulse in this case is purely inductive and should be independent of the magnetic field strength, at least in the range below the first critical field, at which penetration of flux lines into a sample sets in. This type of response is termed a kinetic inductance photoresponse in the literature.

When magnetic vortices are present in a sample, they can affect the impedance of the sample in several ways when it is exposed to short laser pulses.

(a) Magnetic vortices reduce the density of the superconducting condensate by a quantity corresponding to the number of quasiparticles contained in normal vortex cores. In an isotopic superconductor with *s*-wave pairing this results in a magnetic-field dependence of the penetration depth described by the formula $\lambda_L \propto (1-B/B_{c2})^{-1/2}$.

(b) The decrease in the density of the superconducting condensate due to absorption of photons leads to a redistribution of the screening current circulating around a vortex core. The effective radius of the vortex increases, and this leads to a change in the retaining force, which binds the vortex to its pinning center. Since the vortex is continuously acted upon by the Lorentz force generated by the bias current, the abrupt change in the retaining force results in damped oscillations of the vortex about the pinning center. These oscillations contribute to both real and imaginary parts of the sample impedance.

(c) An optical pulse generates a short-time uniform increase in the electron temperature. Since the depinning is caused by activation processes, the temperature rise leads to a larger number of vortices dissociated from their pinning centers per unit time and makes possible local displacements of vortex fragments (magnetic flux creep). The additional mobile vortices or their fragments generated in this process travel in the sample until they are captured by the same or a neighboring pinning center. The thermally activated creep and magnetic flux flow contribute mostly to the real part of the sample impedance variation.

Coffey and Clem[11] suggested the following general expression for the frequency dependent penetration depth taking account of all the mechanisms listed above:

$$\lambda_\omega^2 = \frac{\lambda_L^2 - j\delta_\nu^2/2}{1+2j\lambda_L^2/\delta_{nf}^2}. \tag{4}$$

The effective lengths δ_ν and δ_{nf} characterize the contributions of vortices and unpaired electrons to the impedance,

and λ_L describes the contribution from the superconducting condensate. These three parameters can be expressed by the formulas

$$\lambda_L^2=\frac{\lambda_{L0}^2}{f}, \quad \delta_{nf}^2=\frac{2\rho_n}{\mu_0\omega(1-f)},$$

$$\delta_\nu^2=\frac{2B\Phi_0}{\mu_0\omega\eta}\,\frac{\varepsilon+(\omega\tau_0)^2+j(1-\varepsilon)\omega\tau_0}{1+(\omega\tau_0)^2},$$

$$\tau_0=\frac{\eta}{k_p}\,\frac{I_0^2(\nu)-1}{I_0(\nu)I_1(\nu)}, \quad \varepsilon=\frac{1}{I_0^2(\nu)}, \quad \nu=\frac{U}{2k_BT}, \tag{5}$$

where λ_{L0} is the London penetration depth at zero temperature and magnetic field, Φ_0 is the magnetic flux quantum, ρ_n is the material resistivity in the normal state, η is the viscosity that controls motion of vortices driven by the Lorentz force, k_p is the retaining force per unit length, U is the pinning potential, and I_0 and I_1 are modified Bessel functions of the first kind. The two-liquid approximation yields the following expression for the parameters as functions of magnetic field and temperature:

$$f=\left(1-\left(\frac{T}{T_{c2}}\right)^4\right)\left(1-\frac{B}{B_{c2}(T)}\right),$$

$$B_{c2}(T)=B_{c2}(0)\,\frac{1-(T/T_c)^2}{1+(T/T_c)^2}, \tag{6}$$

$$U=U_0\left(1-\frac{T}{T_{c2}}\right)^{3/2}, \quad k_p=k_{p0}\left[1-\left(\frac{T}{T_{c2}}\right)^2\right]^2,$$

$$T_{c2}=T_c\sqrt{\frac{1+B/B_{c2}(0)}{1-B/B_{c2}(0)}}.$$

Assuming that the pinning potentials and retaining force are determined by the instantaneous relative density of the condensate, we have expressed k_p and U versus temperature in terms of f so as to analyze the vortices' contribution to the photoresponse using a unified formalism. Then the Fourier transform of the electric signal (3) is given by

$$S_\omega=\omega\mu_0 I\,\frac{I}{wd}\,f_\omega\,\frac{d}{df}\{\mathrm{Im}(\lambda_\omega^2)+j\,\mathrm{Re}(\lambda_\omega^2)\}, \tag{7}$$

where f_ω is the Fourier transform of changes in the condensate density as a function of real time. Model calculations have been performed using the following analytic formula for the changes in the condensate density versus time:

$$f(t)=f(0)-\Delta f\,\frac{t^4}{t^4+\tau_{th}^4}\exp\left(-\frac{t}{\tau_f}\right), \tag{8}$$

where the electron thermalization time τ_{th} is the rise time and τ_f is the decay time of the concentration of excess electrons. The electrical signal generated in the sample was expressed using Eqs. (1)–(7). The effect of the electric circuit on the output signal was described using the procedure suggested previously.[6] The calculated signal amplitude was compared to the experimental data. The parameters used in the calculation are listed in Table I.

The simulation of the electric signal revealed that both the shape and amplitude of the calculated response weakly depend on the pinning parameters U and k_p, and the electron thermalization time. In contrast, such parameters as the decay time of the quasiparticle concentration, the London penetration depth, and the relative number Δf of Cooper pairs destroyed by absorption of photons and subsequent multiplication of quasiparticles determine the response to a laser pulse almost completely. The latter two parameters equally affect the positive and negative components of the signal, and have little impact on the shape of the waveform, whereas the decay time affects largely the ratio between the amplitudes of the positive and negative components of the response signal. By varying Δf and λ_L, the calculated amplitude of the positive component was fitted to the experimental data. The best fit was obtained at $\Delta f/f_0=0.05$ and $\lambda_{L0}=1400$ Å. The best coincidence between the shapes of the calculated and recorded signals was achieved (Fig. 1) at a quasiparticle decay time of 3.7 ps, which is in good agreement with investigations of YBaCuO films published earlier.[9,12] The positive signal amplitude versus magnetic field calculated for various limiting cases at a temperature of 50 K is plotted in Fig. 3. Curve a corresponds to the case of immobile vortices ($k_{p0}, U\to\infty$), curve b corresponds to the case with vortex oscillations about pinning centers ($U_0\to\infty$), and curve c shows the calculations that take into account, as compared with the previous case, the flux creep and viscous flow. Note that the calculations plotted by curves a–c were multiplied by a factor of five in order to make the differences among them and the calculations themselves visible in the scale in which the experimental data are plotted. The graph also shows experimental data obtained at various temperatures. It is perfectly obvious that neither the changes in the kinetic inductance taken separately nor the effect of magnetic vortices on the impedance can account for the change of the photoimpedance response measured as a function of magnetic field.

TABLE I. YBaCuO parameters used in simulation.

London penetration depth λ_{L0}	1.4×10^{-7} m
Viscosity of vortex flow η	2×10^{-7} N·s/m
Retaining force per unit length k_{p0}	2.1×10^4 N/m^2
Resistivity in the normal state ρ_n	2.8×10^{-6} Ω·m
Average potential of the pinning center U_0	0.15 eV
Time constant of electron thermalization τ_{th}	900 fs

3.3. Anisotropy of superconducting gap

Since the energy gap in an anisotropic superconductor is more narrow for some directions of the Cooper pair quasimomentum, the critical current density in these directions is also lower. Accordingly, the screening current circulating about a vortex core should lead to electron decoupling at distances considerably larger than the average coherence length. The effective increase in the quasiparticle density caused by this effect was predicted for a superconductor whose gap width tends to zero at some points on the Fermi surface, for example, owing to the *d*-wave symmetry of electron pairing.[13] This result implies that the number of quasiparticles bound to a vortex core is higher by a factor

$\alpha \approx a_\nu/\xi$ in this superconductor in comparison with an isotropic superconductor, where $a_\nu \approx (\Phi_0/B)^{1/2}$ is the average distance between vortices and $\xi(T)=\xi_0/(1-T/T_c)^{1/2}$ is the effective coherence length (ξ_0 is the coherence length at zero temperature). Irrespective of the manner in which vortices move, the band gap anisotropy should increase the amplitude of the impedance response to a laser pulse as a function of magnetic field. We have incorporated the effect of anisotropy by inserting in Eq. (6) the following formula for the superconducting condensate density as a function of temperature and magnetic field:

$$f=\left(1-0.75\frac{T}{T_{c2}}\right)\left[1-\left(\frac{T}{T_{c2}}\right)^4\right]^{1/2}\left[1-\alpha\frac{B}{B_{c2}(T)}\right]. \qquad (9)$$

The temperature dependent component of f in Eq. (9) was constructed using the data[14] on the magnetic field penetration depth in pure YBaCuO crystals. The only fitting parameter used in adjusting the calculations to the experimental data was the coherence length at zero temperature. For each ξ_0 the second critical field was calculated using the expression $B_{c2}=\Phi_0/2\pi\xi_0^2$. The best fit of the calculations to experimental data (Fig. 3) was obtained at a coherence length of 16 Å, which is in reasonable agreement with the coherence length in the *ab* plane measured in YBaCuO by alternative methods.

Note that the signal amplitude as a function of magnetic field is not bound to follow the predictions of the model under discussion, which assumes a uniform distribution of the bias current over the sample cross section and of magnetic flux lines over the sample plane. The configuration of the samples used in the experiments corresponds to the demagnetization factor $\pi d/2(wl)^{1/2} \approx 2\times 10^{-2}$. Therefore, if the first critical field at zero temperature is 500 mT,[10] the onset of vortex penetration in the sample should take place at a magnetic field induction of 2 mT. This magnetic field is generated near sample edges by a current of 20 mA even in the absence of external field. At higher temperatures, the onset of vortex penetration should take place at even lower currents and magnetic fields. As compared with internal regions, sample edges usually contain larger quantities of defects, which can act as pinning centers. If a vortex generated in the sample by the bias current is bound to a pinning center at a distance from the sample edge smaller than λ_L^2/d, the mean density of the screening current flowing between the sample edge and the vortex is higher than that between the vortex and the middle of the sample. The total screening current of such an asymmetric flux line contributes to the response signal, with the polarity of that contribution being the opposite of that due to the bias current that generated the vortex.

An estimate of the mean density of the screening current by the formula $j_m \approx 2\Phi_0 d/\pi\mu_0\lambda_L^4$ indicates that the signal component due to the asymmetric configuration of vortices can have a magnitude equal to that of the contribution from a bias current of one milliampere. An increase in the current or magnetic field leads to a growth in the number of vortices in a sample and their redistribution to the inner region. When the average distance between vortices approaches λ_L^2/d, the signal component due to the asymmetry of vortices vanishes. In our opinion, the suggested model provides a satisfactory explanation of the signal hysteresis in the range of low magnetic fields and the presence of photoresponse at zero bias current.

Taking the normalized condensate density as a function of temperature typical of isotropic (*s*-wave) and anisotropic (*d*-wave) superconductors, we have calculated the signal amplitude versus temperature in zero magnetic field. The comparison between our calculations and experimental data (Fig. 2) shows that calculations by the anisotropic model are in good agreement with experimental data in the ranges of high and intermediate temperatures. The isotropic model, in contrast, provides better agreement with experimental data at $T<0.5T_c$. At lower temperatures, however, the experimental points deviate from both theoretical curves;l specifically, the signal amplitude has a minimum and increases in the range of the lowest temperatures. Since the response amplitude is largely determined by the superconducting condensate density, it is tempting to attribute the increase in the signal amplitude in the range of lower temperatures to a decrease in this density, or equivalently to an increase in the concentration of unpaired particles. Such a behavior was formally derived from measurements of the surface impedance versus temperature using the traditional two-liquid model.[15] Nonetheless, the authors of this result[15] concluded from general considerations that such a description is non-physical and got round the difficulty by assuming that the density of unpaired particles did not tend to zero, as was expected in the case of a pure *d*-wave superconductor, but asymptotically tended to a finite value as the temperature approached zero.

This approach is consistent with a nonvanishing density of quasiparticle states at the Fermi energy, which was predicted for *d*-wave superconductors doped with nonmagnetic impurities.[16] The characteristic energy below which the density of states becomes constant is a function of the impurity concentration. Thus, the linear drop with temperature in the quasiparticle concentration in a dirty *d*-wave superconductor should be replaced in the range of low temperatures by a flatter curve, whereupon the thermal energy of quasiparticles becomes comparable to the characteristic energy.

Since the impurity concentration in the tested samples is unknown, an accurate estimate of the parameters characterizing the effect is impossible. Taking into account impurities in the superconductor, irrespective of their quantity, can lead to better agreement between calculations for a *d*-wave superconductor and measurements, but cannot explain the increase in the signal amplitude at low temperatures. Moreover, our experiments have not revealed that the magnitude of the effect under discussion depends on the sample quality. Although samples irradiated with high-energy gold ions had larger widths of their superconducting transitions, their behavior in the region of low temperatures was almost identical to that of control samples.

It is noteworthy that there is an alternative mechanism that can have a notable effect on the signal amplitude, irrespective of the sample purity. The coefficient of quasiparticle multiplication, in other words, the number of secondary quasiparticles generated in the system by one absorbed photon,

contains contributions from two channels, namely, the electron-electron and electron-phonon ones. The latter yields a higher multiplication coefficient because high-energy electrons release energy needed for dissociation of Cooper pairs through emission of phonons. The lattice anharmonicity, in turn, leads to a lower multiplication mechanism since a phonon emitted by an electron can decay into two phonons with lower energies, which cannot dissociate a Cooper pair and generate additional quasiparticles. This anharmonicity results in a loss of the electron energy and an increase in the bolometric component of the photoresponse. There are several optical phonon modes in YBaCuO, whose anharmonic broadening reduces at lower temperatures.[17] The fraction of the electron energy transmitted to acoustic phonons via these modes can reduce with decreasing temperature, thus increasing the multiplication coefficient and the signal amplitude, accordingly.

4. CONCLUSIONS

We have demonstrated that the increase in the amplitude of the photoimpedance response to optical pulses of YBaCuO caused by external magnetic field cannot be explained in terms of the contribution of magnetic vortices in the mixed state of an isotropic *s*-wave superconductor, but is rather an indication of the anisotropy of the energy gap due to the *d*-wave symmetry of electron pairing. The increase in the signal amplitude at low temperature is tentatively ascribed to the strong anharmonicity of optical phonons in YBaCuO. An accurate description of the temperature dependence of the response amplitude with due account of the quasiparticle density of states due to impurity scattering is possible only in the case of optimized samples with known impurity concentrations.

The author thanks A. V. Sergeev for exceptionally helpful discussions and A. Frenkel for help in conducting experiments. The work was partly financed by the Russian Fund for Fundamental Research and AFOSR research Grant No. F49620-96-1-0225.

*)E-mail: semenov@rpl.mpgu.msk.su

[1] E. Zeldov, N. M. Amer, G. Koren, and A. Gupta, Phys. Rev. B **39**, 9712 (1989).
[2] A. Frenkel, M. A. Saifi, T. Venkatesan, Chilon Lin, X. D. Wu, and A. Inam, Appl. Phys. Lett. **54**, 1594 (1989).
[3] A. M. Kadin, M. Leung, and A. D. Smith, Phys. Rev. Lett. **65**, 3196 (1990).
[4] K. Karraï, E. J. Choi, F. Dunmore, S. Liu *et al.*, Phys. Rev. B **69**, 152 (1992).
[5] M. A. Heusinger, A. D. Semenov, R. S. Nebosis, Yu. P. Gousev, and K. F. Renk, IEEE Trans. Appl. Supercond. **5**, 2595 (1995).
[6] A. D. Semenov, R. S. Nebosis, Yu. P. Gousev, M. A. Heusinger, and K. F. Renk, Phys. Rev. B **52**, 581 (1995).
[7] T. Puig, P. G. Huggard, M. Pont, G. Schneider, J. S. Muñoz, and W. Prettl, Phys. Rev. B **49**, 7004 (1994).
[8] A. V. Sergeev, A. D. Semenov, P. Kouminov, V. Trifonov *et al.*, Phys. Rev. B **49**, 9091 (1994).
[9] M. Lindgren, M. Currie, C. Williams, T. Y. Hsiang, P. M. Fauchet, and R. Sobolevski, J. Sel. Top. Quantum Electronics **2**, 668 (1996).
[10] C. P. Pole, Jr., H. A. Farach, and R. J. Creswick, Superconductivity, Academic Press Inc., San Diego (1995).
[11] M. W. Coffey and J. R. Clem, Phys. Rev. Lett. **67**, 386 (1991).
[12] M. A. Heusinger, A. D. Semenov, Yu. P. Gousev, O. Kus, and K. F. Renk, Inst. Phys. Conf. Ser. **148**, 749 (1995).
[13] G. Volovik, JETP Lett. **58**, 469 (1993).
[14] D. A. Bonn, S. Kamal, K. Zhang, R. Liang *et al.*, Phys. Rev. B **50**, 4051 (1994).
[15] J. R. Waldram, P. Theopistou, A. Porch, and H.-M. Cheah, Phys. Rev. B **55**, 3222 (1997).
[16] L. P. Gor'kov and P. Kalugin, JETP Lett. **41**, 253 (1985); P. A. Lee, Phys. Rev. Lett. **71**, 1887 (1993).
[17] B. Friedl, C. Thomsen, and M. Cardona, Phys. Rev. Lett. **65**, 915 (1990).

Translation provided by the Russian Editorial office

Resonance current–voltage characteristics of three-dimensional tunnel junctions with weak structural disorder

V. Ya. Kirpichenkov*)

Novocherkassk State Technical University, 346400 Novocherkassk, Russia
(Submitted 12 February 1999)
Zh. Éksp. Teor. Fiz. **116**, 1048–1057 (September 1999)

The nonlinear current–voltage characteristics of a three-dimensional tunnel junction with weak (low impurity concentration) structural disorder are obtained for temperature $T=0$, and a formula is derived for the magnitude of the mesoscopic fluctuations in its resonant static tunnel conductance. © *1999 American Institute of Physics.* [S1063-7761(99)02009-0]

1. INTRODUCTION

In this paper some results[1] for a quasi-one-dimensional tunnel junction with weak structural disorder are generalized to the case of a three-dimensional junction in the most interesting (resonant) case, where resonant subbarrier impurity scattering of tunnelling electrons radically alters the current–voltage characteristic of a tunnel junction with impurities compared to that of an ''empty'' (without impurities) tunnel junction. Here we use the concept of quantum resonant-percolation trajectories in tunnel junctions with a weak structural disorder, developed in Ref. 2.

We use the one-electron approximation at $T=0$ and expand in powers of the impurity concentration to determine the form of the nonlinear resonant current–voltage characteristic of a tunnel junction with impurities, to obtain a formula for the magnitude of the mesoscopic fluctuations in its resonant static tunnel conductance, and to find a lower bound estimate for the transverse dimensions of the barrier layer consistent with the condition that these fluctuations be small.

2. MODEL: BASIC EQUATIONS

As in Ref. 1, we shall consider a model of a tunnel junction in the form of an N-I-N sandwich consisting of two identical normal metals N separated by a plane layer of insulator I of thickness L and area S impregnated with impurities.

For the conduction electrons of the N-metals we shall assume a three-dimensional isotropic quadratic dispersion law $\varepsilon=k^2$ ($\hbar^2/2m=1, \hbar=1, m=1/2$) with Fermi energy ε_F.

The electrons in the barrier are assumed not to interact with one another (one-electron approximation), and for the barrier potential $U(\mathbf{r})$ (electronic charge $e=1$) in the region $0\leqslant x\leqslant L$ occupied by the insulator in the absence of an electric field ($v=0$), at the barrier we assume a structural disorder model of the form

$$U(\mathbf{r})=U_0+U_{\text{imp}}(\mathbf{r}), \quad \mathbf{r}=(x,\boldsymbol{\rho}), \quad 0\leqslant x\leqslant L, \tag{1}$$

where $U_0=\text{const}>\varepsilon_F$ is the regular potential of a uniform barrier without impurities, $U_{\text{imp}}(\mathbf{r})$ is a random potential produced by a system of N identical impurities randomly distributed over the insulator layer:

$$U_{\text{imp}}(\mathbf{r})=\sum_{0\leqslant x_j\leqslant L}\hat{u}(\mathbf{r},\mathbf{r}_j), \tag{2}$$

where the points $\mathbf{r}_j$ are macroscopically uniformly distributed over the volume layer $V=SL$ with a density $n=N/V$, and $\hat{u}(\mathbf{r},\mathbf{r}_j)<0$ is the local attractive potential of the electrons to the impurities at the point $\mathbf{r}_j$ with a radius of action r_0.

In the case of low impurity concentrations considered here, the following relationships for the characteristic lengths hold and allow us to use the procedure of expanding the tunnel current in powers of the impurity concentration:[2,3]

$$r_0\ll\alpha_F^{-1}\ll n^{-1/3}<L, \tag{3}$$

where $\alpha=\alpha(\varepsilon)=(U_0-\varepsilon)^{1/2}$ and $\alpha_F^{-1}=\alpha^{-1}(\varepsilon_F)$ is the characteristic damping length for the electronic state with energy ε_F in a uniform barrier.

For voltages $v\ll\varepsilon_F$, $U_0-\varepsilon_F$, and $T=0$, we can write the tunnel current $\langle i(v)\rangle$, tunnel conductance $\langle G(v)\rangle$, and their relative root mean square fluctuations $\langle\delta^2(v)\rangle^{1/2}$ as

$$\langle i(v)\rangle=\int_{\varepsilon_F}^{\varepsilon_F+v}\langle g(\varepsilon)\rangle d\varepsilon, \quad \langle G(v)\rangle=v^{-1}\langle i(v)\rangle, \tag{4}$$

and

$$\langle\delta^2(v)\rangle^{1/2}\left[\frac{\langle i^2(v)\rangle-\langle i(v)\rangle^2}{\langle i(v)\rangle^2}\right]^{1/2}, \tag{5}$$

where

$$\langle i^2(v)\rangle=\int_{\varepsilon_F}^{\varepsilon_F+v}\langle g(\varepsilon)g(\varepsilon')\rangle d\varepsilon\, d\varepsilon', \tag{6}$$

$$g(\varepsilon)\equiv g(\varepsilon,\Gamma_N)=\int\int D(\varepsilon,\mathbf{q},\boldsymbol{\rho},\Gamma_N)\frac{d^2q}{(2\pi)^2}d^2\rho, \tag{7}$$

$D(\varepsilon,\mathbf{q},\boldsymbol{\rho},\Gamma_N)$ is the tunnel transparency of a barrier with a random impurity configuration $\Gamma_N=\{\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N\}$ for electrons with energy ε that have a fixed transverse component of their momentum $\mathbf{q}$ at the entrance to the barrier and a fixed transverse coordinate $\boldsymbol{\rho}$ at the exit, the integral over $\mathbf{q}$ is

taken for all $0 \leqslant q^2 \leqslant \varepsilon$, and that over $\boldsymbol{\rho}$ over the barrier area S. The averages in Eqs. (4) and (6) are taken over the set of impurity configurations $\{\Gamma_N\}$:

$$\langle g(\varepsilon)\rangle = \frac{1}{\Delta\Gamma_N}\int_{\{\Gamma_N\}} g(\varepsilon,\Gamma_N)\,d\Gamma_N, \tag{8}$$

and

$$\langle g(\varepsilon)g(\varepsilon')\rangle = \frac{1}{\Delta\Gamma_N}\int_{\{\Gamma_N\}} g(\varepsilon,\Gamma_N)g(\varepsilon',\Gamma_N)\,d\Gamma_N, \tag{9}$$

where $d\Gamma_N = d\mathbf{r}_1 d\mathbf{r}_2 \cdots d\mathbf{r}_N$, $\Delta\Gamma_N = V^N = (LS)^N$, and $N = nV$.

3. RESONANT TUNNELLING CURRENT AS THE SUM OVER QUANTUM MECHANICAL RESONANCE-PERCOLATION TRAJECTORIES

As the analysis of the quasi-one dimensional case[1] has shown, the most radical difference between the current–voltage characteristic of a barrier with impurities and that for an empty barrier shows up in the conditions for resonant tunnelling, when the energy ε of the tunnelling particles is close to the energy ε_0 of a local single-impurity level. Thus, in the following we shall consider the situation where $\varepsilon_F = \varepsilon_0$ holds for $v = 0$. In this case, for each energy ε close to ε_F the phase space $\{\Gamma_N\}$ factorizes in the form of a set of resonance and nonresonance regions, and the principal contribution to the averages (8) and (9) for the low impurity concentrations considered here is from resonance regions corresponding to isolated weakly twisting quantum mechanical resonant-percolation trajectories.[2]

The calculations of the averages (8) and (9) rely significantly on the following concepts of the spatial structure of the quantum resonant-percolation trajectories for ε close to ε_0. The ideal isolated, shortest, m-center ($m = 1,2,\ldots$) trajectory consists of a strictly periodic chain of m impurities separated by a distance $2y = L/m$ from one another, with the first and last impurities in the chain lying a distance y from the corresponding boundaries from the barrier layer. There should be no other impurities within a tube of radius $\sim 2y$ around the chain, except for those belonging to this trajectory (the isolation condition, which, along with the conditions of chain periodicity and of closeness of ε to ε_0, ensures resonant tunnelling of electrons along quantum resonant-percolation trajectories with a transparency $D_m^{\mathrm{res}} \sim 1$). However, the phase volume $\{\Gamma_N\}$ in the space occupied by such an ideal, strictly periodic quantum resonant-percolation trajectory and, therefore, its probability of formation, are zero. Thus, in calculations using Eqs. (8) and (9), it should be noted that the transmission coefficients along this trajectory, D_m^{res}, do not change significantly and remain on the order of unity, if the coordinates of the impurities differ from those in an ideal quantum resonant-percolation trajectory by amounts $\delta x \lesssim \alpha^{-1}$ along the x-axis and by amounts $\delta\rho \lesssim y\theta$ in the transverse direction (where $\theta \ll 1$ is an angle characterizing the twisting of the trajectory). Here the parameters m, y, θ for weakly twisting trajectories are not independent and are related by

$$m = \frac{\mathscr{L}}{u}\left(1 + \frac{\theta^2}{2}\right), \tag{10}$$

where $\mathscr{L} = \alpha L$ and $u = 2\alpha y$ are, respectively, the dimensionless thickness of the barrier layer and the step size in the quantum resonant-percolation trajectory. Thus, for calculations using Eqs. (8) and (9), any two of the three parameters in Eq. (10) can be chosen as independent. In the following, they are m and u.

Therefore, for $N \gg 1$ and $\alpha_F^3 V \gg 1$ and taking only the principal contribution from the trajectories into account, the average (8) reduces to

$$\langle g(\varepsilon)\rangle = S\sum_{m=1}^{\infty}\int p_m(\varepsilon,u)\,g_m^{\mathrm{res}}(\varepsilon,u)\,du, \tag{11}$$

where

$$p_m(\varepsilon,u) = \alpha^2(\varepsilon)c^m e^{-cm\pi u^3}(u^2\theta^2(m,u))^{m-1} \tag{12}$$

is the probability per unit area barrier layer of forming an isolated m-center quantum resonant-percolation trajectory with a step size u, $c = n\alpha^{-3}$ is the dimensionless impurity concentration, $\theta^2(m,u)$ is expressed through Eq. (10), and

$$g_m^{\mathrm{res}}(\varepsilon,u) = \int\int D_m^{\mathrm{res}}(\varepsilon,\mathbf{q},\boldsymbol{\rho},u)\,\frac{d^2q}{(2\pi)^2}\,d^2\rho, \tag{13}$$

where $D_m^{\mathrm{res}}(\varepsilon,\mathbf{q},\boldsymbol{\rho},u)$ is the transparency of a barrier with a single m-centered trajectory with step size u.

Note that the dependence of $D_m^{\mathrm{res}}(\varepsilon,\mathbf{q},\boldsymbol{\rho},u)$ on v for $v \ll U_0 - \varepsilon_F$, ε_F can be neglected, since it yields a relative correction of order $v/(U_0-\varepsilon_F) \ll 1$ to $D_m^{\mathrm{res}}(\varepsilon,\mathbf{q},\boldsymbol{\rho},u) \sim 1$ calculated for $v = 0$ in a tube of resonant transparency along a quantum resonant-percolation trajectory.

Now, substituting Eq. (11) in Eq. (4), we note that $p_m(\varepsilon,u)$ is a smooth function of ε in the neighborhood of ε_F, while $g_m^{\mathrm{res}}(\varepsilon,u)$ is a rapidly varying function of ε concentrated in the immediate neighborhood of ε_F, so that on integrating with respect to ε it is possible to take $p_m(\varepsilon,u)$ at the point $\varepsilon = \varepsilon_F$ out from under the integral. As a result, we write the tunnel current (4) in the form

$$\langle i(v)\rangle = S\sum_{m=1}^{\infty}\int p_m(\varepsilon_F,u)\,i_m(v,u)\,du, \tag{14}$$

where

$$i_m(v,u) = \int_{\varepsilon_F}^{\varepsilon_F+v} g_m^{\mathrm{res}}(\varepsilon,u)\,d\varepsilon \tag{15}$$

is the tunnel current passing along a single m-center trajectory with step size u.

Thus, Eq. (14) [with Eq. (12)] represents the resonant tunnel current $\langle i(v)\rangle$ in the form of a series in powers of the concentration, whose mth term is the contribution of the m-centered trajectories (to $\langle i(v)\rangle$).

Similarly, Eq. (6) also reduces to

$$\langle i^2(v)\rangle = S^2 \sum_{m,m'=1}^{\infty} \int \left[1+\frac{1}{S}\omega_{m,m'}(\varepsilon_F;u,u')\right] \times p_m(\varepsilon_F,u)p_{m'}(\varepsilon_F,u') \times i_m(v,u)i_{m'}(v,u')du\,du', \tag{16}$$

where

$$\omega_{m,m'}(\varepsilon_F;u,u') = \pi\alpha_F^{-2}u^2\left[\frac{\alpha_F^2}{\pi u^2 p_m(\varepsilon_F,u)}\delta_{m,m'}\delta(u-u')-1\right], \tag{17}$$

$\delta_{m,m'}$ is the Kronecker symbol, and $\delta(u-u')$ is the delta function.

The term $S^{-1}\omega_{m,m'}(\varepsilon_F;u,u')$ in Eq. (16) accounts for the pairwise statistical spatial correlations between trajectories owing to the requirement that they be isolated.

Substituting Eqs. (14) and (16) in Eq. (5), we obtain

$$\langle\delta^2(v)\rangle^{1/2} = \frac{1}{\sqrt{S}}\left[\frac{\Sigma_{m,m'}\int\omega_{m,m'}(\varepsilon_F;u,u')p_m(\varepsilon_F,u)p_{m'}(\varepsilon_F,u')i_m(v,u)i_{m'}(v,u')du\,du'}{\left(\Sigma_m\int p_m(\varepsilon_F,u)i_m(v,u)du\right)^2}\right]^{1/2}. \tag{18}$$

For further calculations using Eqs. (14) and (18), it is necessary to find the tunnel transparency $D_m^{\rm res}(\varepsilon,\mathbf{q},\boldsymbol{\rho},u)$ that appears in Eq. (13).

4. CALCULATING $D_M^{\rm res}(\varepsilon,\mathbf{q},\rho,U)$

The local transparency of a barrier containing an m-center trajectory with a step size u has been found previously[2] for the special case of particles with energy ε normally incident ($\mathbf{q}=0$) on a barrier. Here the problem of calculating the transparency $D_m^{\rm res}(\varepsilon,\mathbf{q},\boldsymbol{\rho},u)$ is generalized to the case of arbitrary $\mathbf{q}$ in the framework of the same method and its dependence on ε is found in more detail.

The Schrödinger equation in the region of a barrier (with one m-center quantum resonant-percolation trajectories) has the form ($v=0$)

$$\Delta\psi-\alpha^2\psi=\sum_{j=1}^{m}\hat{u}_j\psi, \quad 0\leqslant x,x_j\leqslant L,$$
$$\alpha^2=U_0-\varepsilon, \quad \hat{u}_j=\hat{u}(\mathbf{r},\mathbf{r}_j). \tag{19}$$

The continuity conditions are satisfied for ψ and its normal derivatives $\partial\psi/\partial x$ at the barrier boundaries $x=0$ and $x=L$ for all $\boldsymbol{\rho}$.

To the left of the barrier ($x<0$), the function $\psi(x,\boldsymbol{\rho})$ is a superposition of a wave incident on the barrier with transverse momentum $\mathbf{q}$ and a reflected wave:

$$\psi(x,\boldsymbol{\rho}) = a_q\exp(ik_qx+i\mathbf{q}\cdot\boldsymbol{\rho}) + \int b_s\exp(-ik_sx+i\mathbf{s}\cdot\boldsymbol{\rho})\frac{d^2s}{(2\pi)^2},$$
$$k_s=\sqrt{\varepsilon-s^2}. \tag{20}$$

To the right of the barrier ($x>L$) the function $\psi(x,\boldsymbol{\rho})$ is the transmitted wave:

$$\psi(x,\boldsymbol{\rho}) = \int c_s\exp(ik_s(x-L)+i\mathbf{s}\cdot\boldsymbol{\rho})\frac{d^2s}{(2\pi)^2}. \tag{21}$$

The integral with respect to $\mathbf{s}$ in Eqs. (20) and (21) is taken for all $0\leqslant s^2\leqslant\varepsilon$ and the spectral amplitudes c_s and b_s depend, among others, on the trajectory parameters m and u.

The object of the calculation is the transmission coefficient

$$D_m^{\rm res}(\varepsilon,\mathbf{q},\boldsymbol{\rho},u)=\frac{j_x^{\rm out}(\varepsilon,\boldsymbol{\rho},u)|_{x=L}}{j_x^{\rm in}(\varepsilon,\mathbf{q})|_{x=0}}, \tag{22}$$

where

$$j_x^{\rm out}(\varepsilon,\boldsymbol{\rho},u)|_{x=L} = 2{\rm Re}\int k_sc_sc_{s'}^*\exp\{i(\mathbf{s}-\mathbf{s}')\cdot\boldsymbol{\rho}\} \times\frac{d^2s}{(2\pi)^2}\frac{d^2s'}{(2\pi)^2} \tag{23}$$

is the x-component of the density vector of the transmitted flux at the point $(L,\boldsymbol{\rho})$ and

$$j_x^{\rm in}(\varepsilon,\mathbf{q})|_{x=0}=2k_q|a_q|^2 \tag{24}$$

is the x-component of the density vector of the incident flux at the plane $x=0$.

Thus, substituting Eqs. (23) and (24) in Eq. (22), we obtain

$$D_m^{\rm res}(\varepsilon,\mathbf{q},\boldsymbol{\rho},u)={\rm Re}\int\frac{k_sc_sc_{s'}^*}{k_q|a_q|^2}\exp\{i(\mathbf{s}-\mathbf{s}')\boldsymbol{\rho}\} \times\frac{d^2s}{(2\pi)^2}\frac{d^2s'}{(2\pi)^2}. \tag{25}$$

The next problem is to find the relationship between c_s and a_q by solving Eq. (19) with the above boundary conditions at the $x=0$ and $x=L$ planes.

As in Ref. 2, this problem reduces to solving a closed system of m algebraic equations:

$$\varphi_{j+1}-\frac{1}{\mu h}\varphi_j+\varphi_{j-1}=0, \quad 2\leqslant j\leqslant m-1, \tag{26}$$

$$\left(\frac{1}{\mu h}+\frac{\alpha+ik}{\alpha-ik}\right)\varphi_1+\varphi_2=f_q a_q, \tag{27}$$

$$\left(\frac{1}{\mu h}+\frac{\alpha+ik}{\alpha-ik}\right)\varphi_m+\varphi_{m-1}=0, \tag{28}$$

for the quantities

$$\varphi_k=\int \hat{u}(\mathbf{r},\mathbf{r}_k)\psi(\mathbf{r})d\mathbf{r}, \quad k=1,2,\ldots,m, \tag{29}$$

where

$$f_q=-\frac{2ik_q}{\alpha_q-ik_q}\frac{e^{-\alpha_q y}}{4\pi^2 h}, \quad y=\frac{u}{2\alpha},$$

$$h=\frac{\alpha e^{-u}}{4\pi u}, \quad \alpha_q=\sqrt{U_0-\varepsilon-q^2}, \tag{30}$$

and $\mu=\mu(\varepsilon)$ is the amplitude of the subbarrier impurity scattering introduced in Ref. 2.

The continuity condition in the $x=L$ plane implies that the unknown amplitude c_s is given in terms of φ_m (Ref. 2):

$$c_s=-\frac{\exp(-\alpha_s y-i\mathbf{s}\boldsymbol{\rho}_m)}{\alpha_s-ik_s}\varphi_m, \tag{31}$$

where $\boldsymbol{\rho}_m$ is the transverse coordinate of the mth impurity (closest to the plane $x=L$) in the chain.

The solution to Eq. (26) can be written in the form

$$\varphi_j=C_1\lambda_1^j+C_2\lambda_2^j, \quad 2\leqslant j\leqslant m-1, \tag{32}$$

where C_1 and C_2 are constants to be determined and λ_1 and λ_2 are the roots of the characteristic equation

$$\lambda-2\eta\lambda+1=0, \quad \eta=(2\mu h)^{-1}, \tag{33}$$

$$\lambda_{1,2}=\eta\pm i\sqrt{1-\eta^2}. \tag{34}$$

The condition for an energy band with resonant transparency (i.e., the absence of damping φ_j (32) along a quantum resonant-percolation trajectory) is $|\lambda_{1,2}|=1$, which, as can be seen from Eq. (34) is equivalent to requiring $\eta^2\leqslant 1$ or, given Eq. (33) for η, requiring a sufficiently large amplitude of subbarrier scattering,

$$|\mu|\geqslant(2h)^{-1}, \tag{35}$$

which occurs for ε close to ε_0.

Since the subbarrier scattering amplitude for ε close to ε_0 is[2]

$$\mu=\mu(\varepsilon)=-\frac{8\pi\alpha_0}{\varepsilon-\varepsilon_0}, \quad \alpha_0=\sqrt{U_0-\varepsilon_0}, \tag{36}$$

given formula (30) for h we find from Eq. (35) that resonant transmission is possible for

$$|\varepsilon-\varepsilon_0|\leqslant\gamma, \tag{37}$$

where $\gamma=\gamma(u)=4\alpha_0^2 u^{-1}e^{-u}$ and $u=2\alpha_0 y$.

Equations (26)–(28), with Eq. (32), yield a system of equations for C_1 and C_2:

$$\begin{cases} a_{11}C_1+a_{12}C_2=f_q a_q, \\ a_{21}C_1+a_{22}C_2=0, \end{cases} \tag{38}$$

where

$$a_{1j}=\left[\left(\frac{1}{\mu h}+\frac{\alpha+ik}{\alpha-ik}\right)\left(\frac{1}{\mu h}-\lambda_j\right)+1\right]\lambda_j^2,$$

$$a_{2j}=\left[\left(\frac{1}{\mu h}+\frac{\alpha+ik}{\alpha-ik}\right)\left(\frac{1}{\mu h}-\lambda_j^{-1}\right)+1\right]\lambda_j^{m-1}, \quad j=1,2. \tag{39}$$

Having found C_1 and C_2 from Eq. (38), Eqs. (28) and (32) now give

$$\varphi_m=\varphi_m(\varepsilon,\mathbf{q})=\frac{f_q}{\Delta_m(\varepsilon)}\exp[\ln(\lambda_1-\lambda_2)]a_q, \tag{40}$$

where $\Delta_m(\varepsilon)=a_{11}a_{22}-a_{12}a_{21}$ is the determinant of the system (38) calculated for energy ε.

Now, substituting Eq. (40) in Eq. (31), and Eq. (31) in Eq. (25), we find

$$D_m^{\mathrm{res}}(\varepsilon,\mathbf{q},\boldsymbol{\rho},u)=\frac{\alpha^2k^2}{\pi^4(\alpha^2+k^2)^2}\frac{k_q}{k}\times\exp\left\{-\frac{\alpha|\boldsymbol{\rho}-\boldsymbol{\rho}_m|^2}{y}\right\}\times\exp\left\{-\frac{yq^2}{\alpha}\right\}\exp\left\{-\frac{(\varepsilon-\varepsilon_0)^2}{\gamma^2}\right\},$$

$$y=\frac{u}{2\alpha}. \tag{41}$$

It is clear from this that $D_m^{\mathrm{res}}\sim 1$ holds when the following conditions are simultaneously satisfied: $|\boldsymbol{\rho}-\boldsymbol{\rho}_m|<\sqrt{y/\alpha}$, $q<\sqrt{\alpha/y}$, and $|\varepsilon-\varepsilon_0|<\gamma$. Equation (41) is a generalization of the corresponding Eq. (5.16) of Ref. 2 to the case $\mathbf{q}\neq 0$ and refines the dependence of the resonant transparency on ε. For $\mathbf{q}=0$ and $\varepsilon=\varepsilon_0$, these formulas are identical.

5. CURRENT–VOLTAGE CHARACTERISTIC

With Eqs. (41), (13), (15), and (12), the current–voltage characteristic (14) takes the form

$$\langle i(v)\rangle=S\sum_{m=1}^{\infty}\int p_m(\varepsilon_F,u)i_m(v,u)du, \tag{42}$$

where

$$p_m(\varepsilon_F,u)=\alpha_F^2 c^m e^{-cm\pi u^3}[2u^2(mu/\mathscr{S}-1)]^{m-1},$$

$$i_m(v,u)=\frac{1}{8\pi^3\sqrt{\pi}}\frac{k_F^2\alpha_F^2}{(\alpha_F^2+k_F^2)^2}\gamma_F\,\mathrm{erf}\left(\frac{v}{\gamma_F}\right),$$

$$\gamma_F=\gamma_F(u)=4\alpha_F^2 u^{-1}e^{-u},$$

and

$$\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$$

is the probability integral. The function $\langle i(v)\rangle$ is highly nonlinear, and thus differs radically from the corresponding function for an empty barrier.

Formally assuming that the parameter m is continuous, we calculate the right hand side of Eq. (42) by the method of steepest descents. The point of steepest descent is found from the system of equations

$$-3\pi cmu^2+2(m-1)u^{-1}+m(m-1)(mu-\mathscr{L})^{-1}-1-u^{-1}+\xi(u,v)=0,$$
$$\ln c-c\pi u^3+2\ln u+\ln 2+\ln(mu-\mathscr{L})-\ln\mathscr{L}+(m-1)u(mu-\mathscr{L})^{-1}=0, \quad (43)$$

where

$$\xi(u,v)=\frac{\partial}{\partial u}\left[\ln\operatorname{erf}\left(\frac{v}{\gamma_F(u)}\right)\right].$$

The asymptotic solution of the system for $v\sim\gamma_F$, $\mathscr{L}\gg 1$, and $c\mathscr{L}^2\ll 1$ has the form

$$m_0=\mathscr{L}^{1/2}|\ln(c\mathscr{L})|^{-1/2}, \quad u_0=\mathscr{L}/m_0, \quad \theta_0=|\ln(c\mathscr{L})|^{-1/2}. \quad (44)$$

Here we have $\theta_0\ll 1$, which justifies our earlier assumption that the principal contribution to the resonant tunnel current for low impurity concentrations (see Eqs. (47) and (48), below) is from weakly twisting quantum resonant-percolation trajectories.

Let us find the range of concentrations c, within which the resonant tunnel current $\langle i(v)\rangle$ is much greater than the current $i_0(v)$ for an empty barrier

$$\langle i(v)\rangle\gg i_0(v), \quad (45)$$

where

$$i_0(v)=S\frac{4\alpha_F^2k_F^2}{\pi(\alpha_F^2+k_F^2)^2}\frac{\alpha_F^2}{\mathscr{L}}e^{-2\mathscr{L}}v. \quad (46)$$

Substituting Eqs. (42) and (46) in Eq. (45) and limiting ourselves for an estimate to the term with $m=1$ in Eq. (42), we obtain

$$ce^{-c\pi\mathscr{L}^3}\gg\frac{32\pi^2\sqrt{\pi}e^{-2\mathscr{L}}}{\mathscr{L}}\frac{v}{\gamma_F(\mathscr{L})\operatorname{erf}(v/\gamma_F(\mathscr{L}))}. \quad (47)$$

For example, when $v\sim\gamma_F(\mathscr{L})$, for the typical $\mathscr{L}\sim 10$, Eq. (47) gives the estimate

$$10^{-6}\ll c\ll 10^{-2}. \quad (48)$$

Within this range of concentrations, for $\mathscr{L}\sim 10$ and $v\sim\gamma_F(\mathscr{L})$, for the point of steepest descent (44) we find (given that m is a discrete parameter)

$$m_0=1, \quad u_0=\mathscr{L}. \quad (49)$$

Then the current–voltage characteristic takes the form

$$\langle i(v)\rangle=S\alpha_F^2ce^{-c\pi\mathscr{L}^3}i_1(v,\mathscr{L}), \quad (50)$$

where

$$i_1(v,\mathscr{L})=\frac{1}{8\pi^3\sqrt{\pi}}\frac{k_F^2\alpha_F^2}{(\alpha_F^2+k_F^2)^2}\gamma_F(\mathscr{L})\operatorname{erf}\left(\frac{v}{\gamma_F(\mathscr{L})}\right).$$

Under these conditions, for example, with $c\sim 10^{-3}$, the resonant tunnel current $\langle i(v)\rangle$ exceeds $i_0(v)$ by two orders of magnitude.

The differential tunnel conductance

$$G_d(v)=\frac{d\langle i\rangle}{dv}=\frac{S}{4\pi^4}\frac{k_F^2\alpha_F^4}{(\alpha_F^2+k_F^2)^2}ce^{-c\pi\mathscr{L}^3}\times\exp\left[-\frac{v^2}{\gamma_F^2(\mathscr{L})}\right], \quad (51)$$

regarded as a function of v, is a gaussian curve with a characteristic width $\gamma_F(\mathscr{L})$. If, as above, we take $\mathscr{L}\sim 10$, while $\alpha_F^2\sim k_F^2=\varepsilon_F\sim 10$ eV, then the characteristic energy width of the first $(m=1)$ resonance is $\gamma_F(\mathscr{L})\sim 10^{-3}$ eV. This means that a temperature $T\ll 10$ K is needed to observe it experimentally and this resonance itself should show up at voltage scales $v\sim 10^{-3}$ V. Similar estimates can be made for resonances with $m>1$.

Proceeding to a calculation of the mesoscopic fluctuations in the tunnel conductance, we substitute Eq. (17) in Eq. (18) and, given that $p_m^{-1}(\varepsilon_F,u)\gg\pi\alpha_F^{-2}u^2$ holds within the concentration range (48) under consideration, we reduce Eq. (18) to the form

$$\langle\delta^2\rangle^{1/2}=\frac{1}{\sqrt{S}}\left\{\frac{\Sigma_m\int p_m(\varepsilon_F,u)i_m^2(v,u)du}{[\Sigma_m\int p_m(\varepsilon_F,u)i_m(v,u)du]^2}\right\}^{1/2}. \quad (52)$$

Retaining only the principal terms with $m=1$ in the sums in Eq. (52) as an estimate, we obtain

$$\langle\delta^2\rangle^{1/2}=\frac{1}{\alpha_F\sqrt{cS}}\exp\left(\frac{c\pi\mathscr{L}^3}{2}\right). \quad (53)$$

The condition $\langle\delta^2\rangle^{1/2}\ll 1$ yields a lower bound on the area of the tunnel junction

$$\sqrt{S}\gg\frac{1}{\alpha_F\sqrt{c}}\exp\left(\frac{c\pi\mathscr{L}^3}{2}\right), \quad (54)$$

that will ensure real self-averaging of the tunnel conductance for the impurity concentrations considered here.

*)E-mail: p0ncls@novoch.ru

[1] V. Ya. Kirpichenkov, Zh. Éksp. Teor. Fiz. **113**, 1522 (1998) [JETP **86**, 833 (1998)].
[2] I. M. Lifshitz and V. Ya. Kirpichenkov, Zh. Éksp. Teor. Fiz. **77**, 989 (1979) [Sov. Phys. JETP **50**, 499 (1979)].
[3] I. M. Lifshitz, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems*, Pergamon, Oxford (1981) [Russ. orig., Nauka, Moscow (1982)].

Translated by D. H. McNeill

Spin susceptibilities for the Hubbard model in a band approach

A. A. Ovchinnikov and M. Ya. Ovchinnikova*)

Institute of Chemical Physics, Russian Academy of Sciences, 117977 Moscow, Russia
(Submitted 6 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1058–1080 (September 1999)

The properties of spin excitations superposed on a uniform ground state with antiferromagnetic (or spiral) spin structure are studied in a $2D$ Hubbard model. Expressions are derived for the spin susceptibility in the random phase approximation (RPA) using split Hubbard bands as a zeroth approximation. The calculated collective modes with dispersion $\omega(Q)=c|Q-(\pi,\pi)|$ near $Q\sim(\pi,\pi)$ reproduce well the characteristics of the spin excitations observed in undoped cuprates. For doped systems with an antiferromagnetic structure of the ground state, calculating $\chi''(Q,\omega\to 0)$ gives the same mode with a peak at $Q\sim(\pi,\pi)$, regardless of the type of Fermi surface. It is shown that in doped systems with a spiral ground state spin structure, $\chi''(Q,\omega\to 0)$ peaks occur with incommensurate quasimomenta Q that are coupled to the spirality vector. © *1999 American Institute of Physics.* [S1063-7761(99)02109-5]

1. INTRODUCTION

In recent years a number of fundamental discoveries have been made about the electronic structure of the cuprates.[1–3] These include data obtained from angle-resolved photoemission spectroscopy (ARPES)[3] and inelastic neutron scattering,[4] the discovery of the ''small'' Fermi surface[5] and of an anisotropic pseudogap in the ground state of undoped samples,[6,7] etc. Inelastic neutron scattering data occupy a special place. The spin susceptibility $\chi(k\omega)$ has been observed to behave differently in various cuprates as $\omega\to 0$: low-frequency peaks appear in $\chi(k,\omega)$ for $k\sim(\pi,\pi)$ in yttrium ceramic or for incommensurate quasimomenta $k\sim(\pi\pm\delta,\pi)$ in $La_{2-x}Sr_xCuO_4$ (LSCO).[3,8–12] The intensities and dispersion of spin waves have been measured absolutely in doped and undoped cuprates at high frequencies (up to 300 meV).[13–15] Peaks in $\chi(k\omega)$ have been observed in yttrium ceramic that depend on temperature and on the superconducting transition.[16,17]

Many of these results are often treated via a band approach assuming a uniform ground state. For example, according to one interpretation,[18–20] the low-frequency behavior of $\chi(k,\omega)$ in various cuprates is related to the behavior of the Fermi boundary in the unperturbed band. However, all the band theories[18–20] retain a stage with renormalization of the spin susceptibility of the unperturbed system $\chi^0(k\omega)$ to the susceptibilities of a highly correlated system. None of the renormalization variants is entirely clear. For example, in Ref. 18 $\chi^0(k\omega)$ is renormalized by an exchange spin interaction $\sim J\Sigma S_nS_m$. The latter is a consequence of a strong one-center interaction, which changes the band structure significantly. In such a situation, it is impossible to use the unperturbed band, even as a basis. This conclusion follows from the most rigorous calculations of a Fermi liquid with a weak interaction.[21–25] In the case of Fermi boundaries with nesting in the presence of van Hove singularities at four points in the phase plane, the system was shown to be unstable to antiferromagnetic ordering—a spin density wave for $U>0$. However, the perturbation theory used in Refs. 21–25 is not applicable to strong and intermediate coupling, which occurs in the cuprates. Thus, we shall use a medium field variation method, in which the specific instability observed in Ref. 21 is assumed from the start. The corresponding electronic spectrum will be characterized by a band split into two.

The purpose of this paper is to calculate $\chi(k\omega)$, relying on specific variational realizations of the pattern of upper and lower Hubbard bands, to compare the quantitative characteristics of the spin-wave excitations with observations, and to understand whether the interpretation of the incommensurate peaks in the dynamic susceptibility still holds in this approach. This analysis only concerns uniform states with antiferromagnetic and spiral spin structures.

This statement of the problem is of current interest for several reasons.

1. The quantitative characteristics of collective spin excitations are currently known all the way to $\omega\leqslant 300$ meV from absolute inelastic neutron scattering measurements on doped and undoped systems.[13–15] It has become possible to compare these with the characteristics calculated on the basis of a band picture.

2. The idea of splitting of the band into upper and lower Hubbard subbands is fairly well justified for a large U/t ratio and has been used in many approaches.[1,26,27] It is consistent with an explanation of many physical properties of the cuprates in terms of the closeness of the van Hove singularity in the spectrum of states to the Fermi level (the so-called VHS scenario).[2] On the other hand, the Hubbard gap $\Delta_H\sim U$ renormalizes the band gaps, bringing them into agreement with photoemission data, and strengthens the van Hove singularity. However, a two-band picture differs from the VHS scenario of Markievicz.[2,28] According to the latter, the van Hove singularity lies at the center of the band, while the pseudogap $\Delta^*\ll U$ of the ground state of undoped compounds is explained by a small splitting of the band owing to charge density waves or lattice distortions. Another explana-

tion of the pseudogap has been proposed in a picture of split Hubbard bands.[29–32] It is related to the fine details of the upper edge of the lower Hubbard band and to dielectrization of separate parts of the generalized Fermi boundary with low doping. This permits an explanation of the small Fermi surface and the phase diagram in terms of a correlation mechanism for attraction of holes. The latter reduces to a spin–polaron attraction in the $t-J$ model[31] or to its analog, valence-bond correlations in the Hubbard model.[29,30] In each of the models, the d-symmetry superconductivity can be explained without empirical parameters. Here the range of doping for which a two-dimensional antiferromagnetism of the CuO_2 planes exists greatly exceeds the range in which volume antiferromagnetism is observed and overlaps the region in which superconductivity occurs. In both the Hubbard picture[30,31] and the picture of a weakly split single band,[2,28] the type of Fermi surface and the properties of the pseudogap are very sensitive to the small parameter t', which takes the non-nearest neighbors into account, or to the parameter t_{OO}, which characterizes jumps between oxygen atoms. A calculation of the spin excitations of this system should help in making a choice between the two interpretations of the pseudogap Δ^*: either this is a small splitting of a single band[2,28] or it is the energy $\Delta^*(k)=-(E_k-\mu)>0$ of the lower Hubbard band in the dielectric segments of the generalized Fermi boundary.[29,30]

The possibility of describing collective spin perturbations in the band approximation of an unbounded Hartree–Fock method (i.e., an average field method with alternation of the spin projection) has been demonstrated for $1D$ Hubbard models for the electronic structure of the polyenes.[33] In this paper we make an analogous calculation of the spin susceptibility for a $2D$ Hubbard model and discuss its conformity with the observed characteristics of spin wave in the cuprates. In any case, the proposed method is the opposite of the approach of Pines *et al.*[34] In the latter, an empirical spin susceptibility $\chi(k\omega)$ serves to a great extent as the basis for describing the properties of the ground state and, in particular, superconducting coupling. Our goal, on the other hand, is to use explicitly constructed variational functions with different spin structures to calculate the susceptibility, in order to be able to evaluate the reality of spin states with a particular structure.

Our method for calculating $\chi(k\omega)$ is similar to that of Ref. 33, but differs from that of Ref. 35. In the latter, the splitting into lower and upper subbands is described in a Hubbard I approximation. In this approach there is no strict definition of the band states and their structure is not detailed; this shows up in the well-known failure of a number of sum rules. Thus, our results differ substantially from those of Ref. 35. In this paper the properties of the spin excitations are studied for the simplest uniform average-field states, i.e., antiferromagnetic and spiral states, neglecting valence-bond correlations. When the latter are taken into account,[29,30] there is a significant drop in the energy, the dielectric gap is reduced, and there is a narrowing of the doping range within which two-dimensional antiferromagnetism occurs. Thus, valence-bond correlations should also affect the spin excitations of the system. We avoid some topics of current interest, including a description of inhomogeneous charge and spin structures, such as stripe phases with alternation of the antiphase antiferromagnetic domains and concentrations of charge at their boundaries. Rigorous experimental proof of the existence of such static stripe phases has been obtained for $LaNiO_{4+\delta}$,[36,37] $La_{2-x-y}Sr_xNd_yCuO_4$,[38,39] $La_2SrCuO_{4.005}$,[40] and other compounds. Data have been obtained which indicate the presence of analogous dynamic spin fluctuations in $La_{2-x}Sr_xCuO_4$ (LSCO).[40–42] This provides a basis for an alternative interpretation of the incommensurate $\chi(k\omega)$ peaks in LSCO (as opposed to the band interpretation[18–20]).

Despite these remarks, in this paper we shall only consider uniform average-field states, in order to understand how the two-band nature of the Hubbard correlated system affects the spin dynamics.

2. SUSCEPTIBILITY OF THE AVERAGE-FIELD ANTIFERROMAGNETIC STATE

Let us consider a standard Hubbard model which gives a single-band representation of the CuO_2-plane of the HTSC:

$$H=\sum_{k,\sigma}\epsilon_k c^\dagger_{k\sigma}c_{k\sigma}+H_U,\quad H_U=U\sum_n n_{n\uparrow}n_{n\downarrow},$$

$$\epsilon_k=2t(\cos k_x+\cos k_y)+4t'\cos k_x\cos k_y. \tag{1}$$

The spin susceptibility is defined as the Fourier transform of the retarding correlation function:[43,44]

$$\chi_{\alpha\beta}(q\omega)=\langle\langle S^q_\alpha(\omega);S^{-q}_\beta\rangle\rangle = \frac{i}{\hbar}\int e^{-i\omega t}\langle\langle S^q_\alpha(t);S^{-q}_\beta\rangle\rangle dt, \tag{2}$$

where

$$\langle\langle A(t);B\rangle\rangle=\begin{cases}\langle[A(t),B]\rangle, & t>0,\\ 0, & t<0.\end{cases} \tag{3}$$

In Eq. (2), the q-components of the spin and density $\rho^q=0.5n^q$ can be written in a uniform fashion:

$$S^q_\alpha=\frac{1}{2}\sum_{ss'}(\sigma_\alpha)_{ss'}r^q_{ss'},\quad \rho^q=\frac{1}{2}\sum_{ss'}(\sigma_0)_{ss'}r^q_{ss'}, \tag{4}$$

where

$$r^q_{ss'}=\frac{1}{\sqrt{N}}\sum_n e^{iqn}\langle c^\dagger{}_{ns}c_{ns'}\rangle. \tag{5}$$

Here the σ_α with $\alpha=1,2,3$ (or x,y,z) are the Pauli matrices and σ_0 is the unit matrix.

In the absence of an interaction ($U=0$), the susceptibility tensor is isotropic,

$$\chi^0_{\alpha\beta}(q\omega)=\delta_{\alpha\beta}\frac{1}{2}\chi^0(q\omega),$$

$$\chi^0(q\omega)=-\frac{1}{N}\sum_k\frac{f_{k+q}-f_k}{\epsilon_{k+q}-\epsilon_k-\omega+i\gamma} \tag{6}$$

and is determined by the band energies ϵ_k and the corresponding Fermi functions $f_k = f(\epsilon_k/kT)$. The same sort of expressions exist for the density correlation when $U=0$.

There are several equivalent methods for deriving the susceptibilities, renormalized by the interaction: the random phase approach (RPA), factoring the equations of motion, and summing ladder diagrams.[44] All of these yield the same result:

$$\chi_{\alpha\alpha}(q\omega) = \frac{1}{2}\frac{\chi^0(q\omega)}{1-U\chi^0(q\omega)}, \quad \alpha = x,y,z. \tag{7}$$

For large U/t, this approach cannot be satisfactory in the case of a half-filled band or lightly doped systems, since their spectra are far from the zero spectrum of ϵ_k. For these systems, the lowest states of the average field are antiferromagnetic (or spiral) states, in which the one-electron levels form upper and lower Hubbard bands, separated by a gap

$$E_{k\lambda=1(2)} = \frac{\epsilon_k + \tilde{\epsilon_k}}{2} \mp \sqrt{\delta\epsilon_k^2 + U^2 d_0^2},$$

$$\delta\epsilon_k = \frac{1}{2}(\epsilon_k - \tilde{\epsilon_k}), \quad \tilde{k} = k - (\pi,\pi). \tag{8}$$

Here $d_0 = (-1)^n \sigma/|\sigma|\langle n_{n\sigma}\rangle$ is the alternating spin density.

An extension of these studies to correlated states has shown[29,30] that when valence-bond correlations are taken into account, the two-dimensional antiferromagnetism and splitting of the bands into two subbands still occur over a fairly wide range of doping. The most convincing argument in favor of this picture is a determination that a pseudogap in the normal state is related to the dielectrization of individual parts of the Fermi boundary and is possible only when a gap exists between the upper and lower subbands.[29,30] This means that the spectrum (8) of a real split band must be used for calculating the null susceptibility $\chi^0(q\omega)$ and for its subsequent renormalization taking the interaction into account. Carrying out this program is the purpose of this paper.

For simplicity we take the antiferromagnetic solutions of the average field neglecting valence-bond correlations as a basis (although, according to Refs. 29 and 30, it is precisely the latter which ensures attraction of holes in the d-channel and superconductivity). Thus, we divide the initial Hamiltonian (1) into a linearized Hamiltonian corresponding to a self-consistent solution with alternating spins and the rest, which appears as a perturbation

$$H = H_L + V, \quad V = H_U - (H_U)_L. \tag{9}$$

The eigenstates of the linearized Hamiltonian corresponding to the band energies (8) are

$$b^\dagger_{k\lambda\sigma} = \sum_i \{c^\dagger_{k\sigma}, c^\dagger_{\tilde{k}\sigma}\}_i U_{i\lambda}(k\sigma),$$

$$i,\lambda = 1,2, \quad \tilde{k} = k + (\pi,\pi), \tag{10}$$

where

$$U_{i\lambda}(k\sigma) = \begin{pmatrix} \cos\varphi_k & -\xi_\sigma \sin\varphi_k \\ \xi_\sigma \sin\varphi_k & \cos\varphi_k \end{pmatrix},$$

$$\tan 2\varphi_k = \frac{\delta\epsilon_k}{Ud_0}, \quad \xi_\sigma = \frac{\sigma}{|\sigma|}, \quad \delta\epsilon_k = \frac{1}{2}(\epsilon_k - \tilde{\epsilon_k}). \tag{11}$$

The inverse transformation of the basis operators is written

$$c^\dagger_{k\sigma} = \sum_\lambda U_{i_k\lambda}(\underline{k}\sigma) b^\dagger_{\underline{k}\lambda\sigma}, \tag{12}$$

where $\underline{k}$ is the quasimomentum introduced into the Brillouin magnetic zone F ($|\underline{k}_x \pm \underline{k}_y| < \pi$) and i_k is the corresponding reduction index: $i_k = 1$, $\underline{k} = k$ for $k \in F$ or $i_k = 2$, $\underline{k} = k - (\pi,\pi)$ for $k \notin F$.

The self-consistency condition for the solution, i.e., the equation for d_0, has the familiar form

$$1 = U\frac{1}{N}\sum_k^F \frac{1}{2g_k}(f_{k1} - f_{k2}), \quad g_k = \sqrt{\delta\epsilon_k^2 + U^2 d_0^2}. \tag{13}$$

The index F on the summation sign means that the sum over k is taken within the Brillouin zone and the $f_{k1(2)}$ are the Fermi functions for the states of the lower and upper bands.

This equation gives a high value for the critical doping level δ_c corresponding to a transition of the antiferromagnetic solution into a paramagnetic one: $\delta_c \sim 0.45$ for $U/t = 8$. When the valence-bond correlations are taken into account[29,30] this level falls to $\delta_c \sim 0.3$. However, both values exceed the region where antiferromagnetism exists ($\delta_{\text{exp}} \sim 0.05$). Nevertheless, variational calculations of the energy[30,45] and calculations by the slave-boson method[46,47] suggest a large radius, $R_{AF} \gg a$, for the two-dimensional antiferromagnetic correlations. We note also that the correct Néel temperature $T_N \sim 4t^2/U$ cannot be obtained from Eq. (13), which would imply $T_N \sim U/4$ (for $U/t \sim 8$). T_N may decrease when multiple electron scattering is taken into account using parquet diagrams.[21–23] But, again, we emphasize that this sum is meaningful only for small parameters U/t. For large U/t the impossibility of describing T_N using Eq. (13) means, most likely, that it is necessary to proceed to inhomogeneous solutions. This follows from the idea[2,37] that the long-range antiferromagnetic ordering breaks down when dynamic or static disordered antiferromagnetic domains or local spiral states develop. Proof of local magnetic ordering in the cuprates over a wide range of doping has been provided by data on the nuclear quadrupole resonance of the copper nuclei, μ-meson spin resonance, and Mössbauer measurements.[2]

Because of the doubling of the unit cell among the correlators $\langle\langle r^q(\omega); r^{-q'}\rangle\rangle$, not only the correlators with $q' = q$ but also those with $q' = q \pm (\pi,\pi) = \tilde{q}$ will be nonzero. We shall now characterize the complete set of quasimomenta of the Brillouin zones G of the initial lattice by the vector $\underline{q}$ normalized to the Brillouin magnetic zone F and the normalization indices $i_q = 1,2$:

$$\underline{q} = q + (\pi,\pi)(i_q - 1), \quad \underline{q} \in F. \tag{14}$$

Thus, we determine the entire set of spatial harmonics (5) for the spin and density as

$$r^q_{\sigma\sigma'} = Y^{q,1}_{\sigma\sigma'}, \quad r^{\tilde{q}}_{\sigma\sigma'} = Y^{q,2}_{\sigma\sigma'}, \quad q \in F, \quad \tilde{q} = q + (\pi,\pi). \tag{15}$$

The expression for these components in the basis of zone operators $b_{k\sigma}$ takes the form

$$Y^{q,m}_{\sigma,\sigma'}=\frac{1}{N}\sum_{k}^{F}\sum_{\lambda\lambda'}M^{m}_{\lambda\lambda'}(qk\sigma\sigma')b^{\dagger}_{\underline{k+q}\lambda\sigma}b_{k\lambda'\sigma'}, \quad (16)$$

where

$$M^{m}_{\lambda\lambda'}(qk\sigma\sigma')=\sum_{ij}U_{i\lambda}(\underline{k+q},\sigma)U_{j\lambda'}(k,\sigma') \times\delta^{(2)}(i+j+m+i_{kq}). \quad (17)$$

Here $\lambda,\lambda',i,j=1,2$ and the vector $k+q=k+q+(\pi,\pi)(i_{kq}-1)$ is normalized to the Brillouin magnetic zone with the corresponding normalization index i_{kq} and $\delta^{(2)}$ is δ *modulo* 2.

The symmetry of the problem implies that there is no correlation $\langle\langle Y^{q,m}_{\sigma,\sigma'};Y^{-q,m'}_{s,s'}\rangle\rangle$ between the transverse ($\sigma=-\sigma'$) and longitudinal ($\sigma=\sigma'$) components. We begin by calculating the nonzero transverse correlators

$$D_{ij}(q\omega\underline{\sigma}\sigma)=\langle\langle Y^{q,i}_{\underline{\sigma},\sigma};Y^{-q,j}_{\sigma,\underline{\sigma}}\rangle\rangle. \quad (18)$$

Of the equivalent methods of deriving an expression for the susceptibility, we shall use the simplest, factoring the equation of motion.[43] Turning to the representation (16) for Y^{qi}, we consider the correlation functions

$$A^{qj}_{\lambda_1\lambda_2}(k\underline{\sigma}\sigma t)=\langle\langle\vartheta^{q}_{\lambda_1\lambda_2}(\underline{\sigma},\sigma,t);Y^{-q,j}_{\sigma,\underline{\sigma}}\rangle\rangle, \quad (19)$$

where $\underline{\sigma}=-\sigma$ and the operator

$$\vartheta^{q}_{\lambda\lambda'}(\underline{\sigma},\sigma)=b^{\dagger}_{\underline{k+q}\lambda\underline{\sigma}}b_{k\lambda'\sigma} \quad (20)$$

is taken at time t. Following Ref. 43 and differentiating Eq. (20) with respect to time, we obtain

$$i\frac{d}{dt}A^{qj}_{\lambda\lambda'}(k\underline{\sigma}\sigma t)=\delta(t)\langle[\vartheta^{q}_{\lambda\lambda'}(\underline{\sigma},\sigma),Y^{-q,j}_{\sigma,\underline{\sigma}}]\rangle|_{t=0} +(E_{\underline{k+q}\lambda}-E_{k\lambda'})A^{qj}_{\lambda\lambda'}(k\underline{\sigma}\sigma t) +\langle\langle[\vartheta^{q}_{\lambda\lambda'},V]_{t};Y^{-q,j}_{\sigma,\underline{\sigma}}\rangle\rangle. \quad (21)$$

Here $E_{k\lambda}$ is the energy (8) of the split bands ($\lambda=1,2$) and an expression for the perturbation V in terms of the operators $b_{k\lambda\sigma}$ is given in the Appendix, Eqs. (A1)–(A3). For calculating the last term in Eq. (21), in the commutator $[\vartheta,V]$ we retain only those operators which reproduce a particle–hole pair $b^{\dagger}_{\underline{k'+q}\lambda_1\underline{\sigma}}b_{k'\lambda_2\sigma}$ with the same total momentum q against the background of the ground antiferromagnetic state of the average field (the approximation of a low excitation concentration). As a result, we obtain

$$[\vartheta^{q}_{\lambda\lambda'},V]|_{t}\simeq-U(f_{\underline{k+q},\lambda}-f_{k\lambda'}) \times\sum_{i}M^{i}_{\lambda\lambda'}(qk\underline{\sigma}\sigma)Y^{q,i}_{\underline{\sigma},\sigma}(t). \quad (22)$$

Here the operators Y and functions M are given by Eqs. (16) and (17), while all the quasimomenta are normalized, i.e., $k, q, k+q\in F$. The Fermi functions $f_{k\lambda}$ correspond to the levels $\overline{E_{k\lambda}}$ of the upper and lower Hubbard bands. Repeating the steps of the derivation in Ref. 43, we find expressions for the Fourier transforms $A^{qj}_{\lambda_1\lambda_2}(k\underline{\sigma}\sigma\omega)$ and $D_{ij}(q\omega\underline{\sigma}\sigma)$. The latter, in turn, are expressed in terms of a sum over k' of the $A^{qj}_{\lambda_1\lambda_2}(k'\underline{\sigma}\sigma\omega)$ with the corresponding weighting functions (17). As a result, we arrive at a system of algebraic equations for the correlators D_{ij} [see Eqs. (18) and (19)]. The solution of these equations gives a matrix analog of the renormalized transverse susceptibility for the case of split bands:

$$D_{ij}(q\omega\underline{\sigma}\sigma)=\{[\hat{I}-U\hat{D}^{0}]^{-1}\hat{D}^{0}\}_{ij}. \quad (23)$$

The matrix of second rank, D^{0}_{ij}, in Eq. (23) is given by

$$D^{0}_{ij}=-\frac{1}{N}\sum_{\lambda\lambda'}\sum_{k}^{F}\frac{f_{\underline{k+q}\lambda}-f_{k\lambda'}}{E_{\underline{k+q}\lambda}-E_{k\lambda'}-\hbar\omega+i\gamma}R^{ij}_{\lambda\lambda'}, \quad (24)$$

with

$$R^{ij}_{\lambda\lambda'}(q,k,\underline{\sigma}\sigma)=M^{i}_{\lambda\lambda'}(qk\underline{\sigma}\sigma)M^{j}_{\lambda\lambda'}(qk\underline{\sigma}\sigma). \quad (25)$$

Using the definitions (17) and (11), we obtain the following compact expressions for R^{ij}:

$$R^{ii}_{\lambda\lambda'}(q,k,\underline{\sigma}\sigma)=\frac{1}{2}\left\{1+(-1)^{\lambda-\lambda'} \times\left[(-1)^{i+i_{kq}}\frac{\delta\epsilon_{kq}\delta\epsilon_{k}}{g_{kq}g_{k}}-\frac{\Delta^{2}}{g_{kq}g_{k}}\right]\right\}, \quad (26)$$

and

$$R^{12}_{\lambda\lambda'}(q,k,\underline{\sigma}\sigma)=R^{21}_{\lambda\lambda'}(q,k,\underline{\sigma}\sigma) =\frac{-\sigma}{2|\sigma|}\left\{(-1)^{\lambda'}\frac{\Delta}{g_{kq}}-(-1)^{\lambda}\frac{\Delta}{g_{k}}\right\}, \quad (27)$$

where

$$\delta\epsilon_{k}=\frac{1}{2}(\epsilon_{k}-\epsilon_{\tilde{k}}),\quad g_{k}=\sqrt{\delta\epsilon_{k}^{2}+\Delta^{2}},$$

$$\tilde{k}=k+(\pi,\pi),\quad \Delta=Ud_{0}, \quad (28)$$

while $\delta\epsilon_{kq}$ and g_{kq} are analogous functions of the reduced quasimomentum $k+q=k+q+(\pi,\pi)(i_{kq}-1)$, where $i_{kq}=1$ (2) for $k+q\in\overline{F}(\notin F)$.

The diagonal elements of the matrix D_{ij} calculated using Eq. (24) directly determine the transverse susceptibility

$$\chi''_{xx}(Q\omega)=\chi''_{yy}(Q\omega)=\frac{1}{2}\chi''_{-+}(Q\omega) =\frac{1}{2}\,\mathrm{Im}\{D_{i_Q i_Q}(q\omega\underline{\sigma}\sigma)\}, \quad (29)$$

where $q=Q-(\pi,\pi)(i_Q-1)$ is the quasimomentum normalized to the magnetic Brillouin cell and the index $i_Q=1$ (2) for $Q\in F(\notin F)$. Thus, the diagonal components D_{ii} are directly observed in inelastic neutron scattering. The nondiagonal components D_{ij} ($j\neq i$) can only contribute to the local characteristics of the spin system, in particular, those measured in NMR experiments.

For the paramagnetic state of the average field ($\Delta=0$, $U\neq0$), the diamagnetic components of the matrix (23) transform to the well known expressions (7) and (6) for the renormalized susceptibilities $\chi_{+-}(Q\omega)$, calculated using an

unsplit band as a zeroth approximation. For the undoped antiferromagnetic case $n=1$ for the Fermi functions in Eq. (24) we have $f_{k1}=1$ and $f_{k2}=0$. That is, the low-frequency, collective spin–wave perturbations are determined in this case by an interband particle–hole interaction. The dispersion $\omega(q)$ of the transverse spin wave is found from the equation for the pole in Eq. (24), i.e., from the equation

$$\mathrm{Det}=\|\delta_{ij}-UD^0_{ij}(q\omega_{\bar{\sigma}\sigma})\|=0. \tag{30}$$

We shall show that for $q\to 0$ and $\omega\to 0$, the determinant behaves as

$$\mathrm{Det}=aq^2-b\omega^2+O(q^3,\ldots). \tag{31}$$

In fact, for $q\to 0$ and $\omega\to 0$ we have $D_{12}\sim\omega$, $1-UD_{11}\to\mathrm{const}+O(q^2,\omega^2)$, and

$$1-UD^0_{22}=\left[1-\frac{U}{N}\sum_k{}^F\frac{f_{qk,1}-f_{k2}}{2g_k}\right]+(aq^2-b'\omega^2)+O(q^3). \tag{32}$$

But the expression in square brackets in Eq. (32) goes to zero according to Eq. (13), which defines the self-consistent quantity d_0 for alternating spin. As a result, for Q near (π,π) or $|q|=|Q-(\pi,\pi)|\ll\pi$, we have

$$\chi''_{-+}=\mathrm{Im}\,D_{22}(q\omega)=A\delta(c^2q^2-\omega^2)=\frac{A}{2cq}\delta(\omega-c|q|),\quad q<1. \tag{33}$$

Therefore, the band calculation of the spin susceptibility with Hubbard splitting of the bands reproduces the characteristic dispersion of transverse spin waves, $\omega(Q)=c|q|$, for Q near (π,π) obtained in the Heisenberg model for an undoped antiferromagnet[44] and observed distinctly in a number of cuprates.[13–15] Similar algebraic equations (see the Appendix) describe the correlation of the density and longitudinal susceptibility, χ_{zz}, where z is the axis of spin quantization of the antiferromagnetic state. As opposed to the transverse components, there is no low-frequency collective mode of the spin excitations with z-polarization.

To conclude Sec. 2 we introduce an example of a model that characterizes the gap in the spin excitation spectrum. We supplement the Hamiltonian (1) with the interaction $\Delta H=B_0(-1)^n S_{zn}$ of the spins with an alternating ''magnetic field'' parallel to the axis of spin quantization of the antiferromagnetic state. This interaction can serve as a crude model of the spin exchange interaction of neighboring CuO_2 planes. Then the band energy spectrum is given by Eq. (8), but with a new $g_k=\sqrt{(\delta\epsilon_k^2+(Ud_0-B_0)^2}$, while the self-consistent value of d_0 of the alternating spin is now found from the equation

$$1=U\frac{1}{N}\sum_k{}^F\frac{1}{2g_k}\left(1-\frac{B_0}{Ud_0}\right)(f_{k1}-f_{k2}) \tag{34}$$

with the new g_k. Equation (34) has two solutions with different signs, $\nu=\mathrm{sign}(d_0B_0)$. For the solution with $\nu=-1$, which corresponds to the lower energy, at low ω/t, B_0/t, and $q=|Q-(\pi,\pi)|<1$, the quantity $1-UD^0_{22}$ has the following expansion:

$$1-UD^0_{22}=(aq^2-b'\omega^2-\nu\Delta^2_{Sp})+O(q^3). \tag{35}$$

It differs from Eq. (32) in having a spin gap term $\Delta_{Sp}=|B_0/(Ud_0)|$. As a result, for the lowest (in terms of energy) antiferromagnetic state ($\nu=-1$) Eq. (35) we obtain a typical spin-gap dependence for the excitation spectrum:

$$\omega(Q)=\sqrt{c^2q^2+\Delta^2_{sp}},\quad q=Q-(\pi,\pi). \tag{36}$$

For the metastable antiferromagnetic solution with $\nu=1$, on the other hand, $\omega(Q)$ goes to zero for finite $|q|\neq 0$, which indicates an unstable solution. In this model a gap shows up for any polarization of the transverse spin waves. In a number of cuprates[14,15] a gap is observed for just one of the components of the transverse spin excitations, with a polarization perpendicular to the ab-plane. One possible description of this behavior might be provided by a model that introduces an alternating magnetic field perpendicular to the orientation of the average spins of the antiferromagnetic state. This kind of model is consistent with a noncollinear arrangement[48] of the antiferromagnetic alternating spins of the different CuO_2 layers in Pr_2CuO_4. However, a quantitative description requires introducing an anisotropy in the spin interactions and fixing the antiferromagnetic axis in space. A complete accounting for this type of spin–orbital and dipole-spin–spin interactions is a separate problem.

3. DISCUSSION OF RESULTS: ANTIFERROMAGNETIC STATES

We begin with undoped systems ($n=1$). In this case, the completely filled lower Hubbard band ($f_{1k}=1$) is separated by a gap from the unfilled upper band ($f_{2k}=0$) and the spin excitations are determined solely by ''interband'' particle–hole pairs. Here the two-particle excitations $\omega_q^{(0)}\geqslant 2\Delta$ [the poles of the ''zero'' susceptibility (23)] are characterized by the gap 2Δ, so that $\mathrm{Im}\,D_{ij}^{(0)}=0$ holds for $\omega<2\Delta$, $\gamma\to 0$. The dispersion of the low-frequency branch of the collective excitations ω_q is found from an equation for the pole of the renormalized susceptibility (24) as $\gamma\to 0$, while the intensity $I(Q)$ in the expression

$$\chi''_{x(y)}(Q\omega)=I(Q)\delta(\omega-\omega_q),\quad q=Q-(\pi,\pi)i_Q\in F \tag{37}$$

is determined minus the function $D_{ii}(\omega)$ as $\gamma\to 0$. A clear representation is provided by directly calculating χ'' according to Eqs. (33) and (23) for finite $\gamma>0$, which simulates a finite energy resolution. Figure 1 shows a typical $\chi''(Q,\omega)$ curve for a series of values of ω along the $q_x=q_y$ cross section for $\gamma=0.008t$ (q is the quasimomentum normalized to the magnetic Brillouin zone). The location of the peak repeats the dispersion ω_q found from Eq. (30). The peak height and its width $\Delta\omega_{1/2}$ depend on the artificially introduced parameter γ in such a way that the peak intensity

$$I(Q)=\int\chi''(Q\omega)d\omega=\pi\gamma\Delta\omega_{1/2}$$

is independent of γ for small γ. Figure 2 is a typical plot of ω_Q for Q varying along a $\Gamma MY\Gamma$ contour. The function ω_Q is periodic within the Brillouin magnetic zone, i.e., ω_Q near the point $Y(\pi,\pi)$ repeats ω_Q near the point $\Gamma(0,0)$. As op-

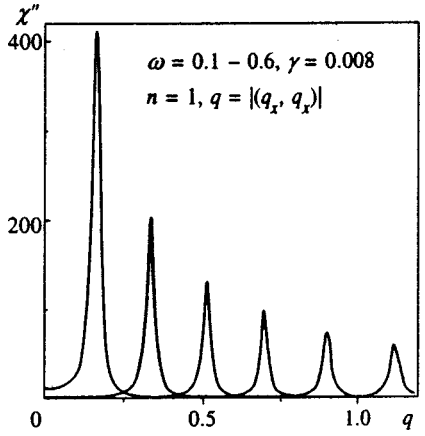

FIG. 1. The susceptibilities $\chi''(q,\omega)$ as functions of the reduced quasimomentum q (in units of rad/a) for $\omega=0.1$–0.6 in an undoped system with $U=8$ and $t'=-0.05$. The curves were calculated for $\gamma=0.008$. U,t',ω, and γ are all given in units of t.

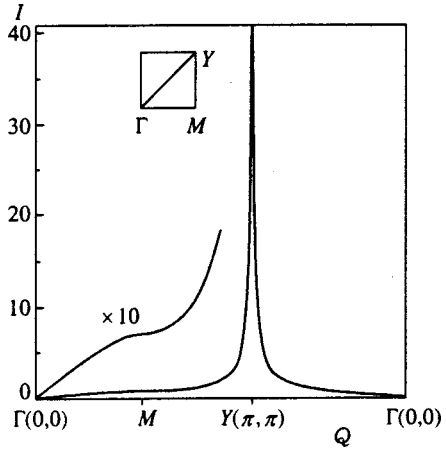

FIG. 3. The intensity $I(Q)$ of the peak in $\chi''(Q,\omega)$ given by Eq. (37) for quasimomentum that varies along the contour $\Gamma-M-Y-\Gamma$. The model parameters are as in Fig. 2.

posed to ω_Q, the intensity $I(Q)$ is periodic only in the ground Brillouin zone. Figure 3 shows a plot of the dimensionless $I(Q)$ for quasimomentum Q varying along a $\Gamma MY\Gamma$ contour. The function $I(Q)$ behaves differently, $\propto|Q|$ or $\propto|Q-(\pi,\pi)|^{-1}$, respectively, in the neighborhoods of the points Γ and Y. Thus, for an undoped dielectric, the band calculation, like the standard description of spin waves in the Heisenberg model, yields a linear dispersion $\omega_Q\sim cq$ for low ω and the absorption peak corresponds to $Q=(\pi,\pi)$ in accordance with the dependence $I(Q)\propto 1/q$ for $q=|Q-(\pi,\pi)|<1$.

We now compare the calculated characteristics of the spin waves of the undoped systems with experiment. The spin wave velocity $c=d\omega/dq|_{q=0}$ varied over $c=0.60$–$0.58\,t$/rad in our models with $U/t=8$ and $t'=\pm 0.05t$. For the estimated[49,50] $t=0.4$–0.5 eV ($J=4t^2/U=80$–125 meV), we obtain $c=880$–1140 meV·Å. These values are entirely consistent with the values $c_{\rm exp}=850$, 1020, 800 meV·Å measured in the cuprates La_2CuO_4, Nd_2CuO_4, and Pr_2CuO_4, respectively.[13] Absolute measurements[13] of the inelastic scattering cross sections in the high-frequency region yielded a spectral weighting characteristic, specifically

$$\tilde{\chi}_{2D}(\omega)=\frac{1}{2}(g\mu_B)^2\int dQ_{2D}\,\mathrm{Im}\,\chi(Q\omega)\Big/\int dQ_{2D}. \tag{38}$$

The factor containing the Bohr magneton μ_B in front of the integral is a consequence of the different definitions of χ in Eq. (2) and in Ref. 13. In the experimental region of $\omega\leqslant 300$ meV, Eq. (38), which is almost independent of ω, gave $\tilde{\chi}_{\rm exp}=2.7$, 1.8, 2.3 μ_B^2/eV for the same cuprates.[13] The value of $\tilde{\chi}_{2D}$ that we have calculated for a model with $U/t=8$ also depends weakly on ω (see the inset to Fig. 4) and is given by $\tilde{\chi}_{2D}\approx 2\,\mu_B^2/t$, which equals $\tilde{\chi}_{2D}\approx 5$–$4\,\mu_B^2$/eV for $t=0.4$–0.5 eV. As expected, this result is less than that,

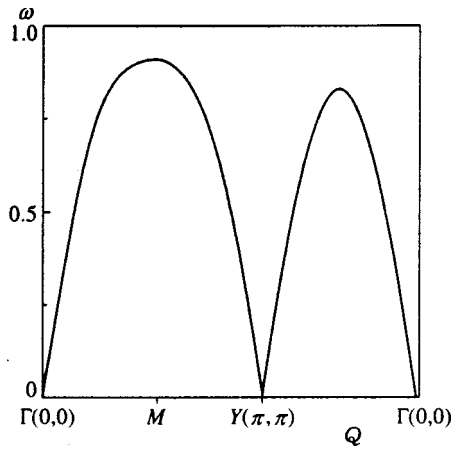

FIG. 2. Dispersion of the collective spin excitation mode $\omega(Q)$ (in units of t) for quasimomentum varying along the contour $\Gamma(0,0)-M(\pi,0)-Y(\pi,\pi)-\Gamma$. The model parameters are $U/t=8$, $t'/t=-0.05$, and $n=1$.

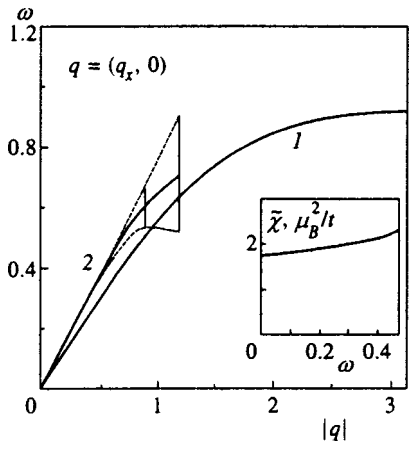

FIG. 4. The frequency $\omega(q)$ of the collective mode (in units of t) as a function of the reduced quasimomentum $q=|(q_x,0)|$ (in radians) for an undoped system (curve *1*) and for doping $1-n=0.15$ (curve *2*). Curve 2 was calculated on the basis of the antiferromagnetic solution for the average field as the dependence of the position of the maximum, $\omega_{\max}(q)$, of the susceptibility $\chi''(q,\omega)$ for fixed q. The length of the vertical segments equals the half width $\Delta\omega_{1/2}$ of the peak. The dashes are for a clear representation of the width of the mode. The inset shows the dependence of Eq. (38) on ω in units of μ_B^2/t.

$\tilde{\chi}_{2D}^{cl}=S(g\mu_B)^2/2J$ ($S=1/2$, $J=4t^2/U$) obtained by the classical Heisenberg model neglecting quantum-mechanical corrections. This last value is $\tilde{\chi}_{2D}^{cl}=8\,\mu_B^2/eV$ for $t=0.5$ eV ($J=125$ meV). Our calculated values are twice the experimental results. We believe the main reason[13] for the discrepancy is the following. During the comparison with experiment, the factor α_d^4, which reflects the amplitude of the α_d d-orbit of copper in the hybridized $p-d$ orbital (single-hole ''node'' orbit of the single-band Hubbard model[49,50]), was left out of the calculated form factor for neutron scattering on Cu. For the estimated $\alpha_d\sim 0.86$ (Refs. 49 and 50) a correction factor of $\alpha_d^4\approx 0.55$ makes it possible to match the calculated value of $\tilde{\chi}$ with experiment.

Thus, the band approach yields a reasonable quantitative description of spin excitations in undoped systems. This means that it can be applied to doped systems in order to answer three questions: (1) how does doping affect the dispersion of spin waves and relaxation (the width of the absorption peaks)? (2) Does the change in the type of Fermi surface associated with the sign of t' in Refs. 29–31 actually change the low-frequency behavior of the susceptibility $\chi''(Q,\omega)$ in a fundamental way? 3) How do the model parameters influence the characteristics of spin waves?

Calculations of $\chi(Q\omega)$ for doped systems ($n-1\leqslant 0.2$) show that for low ω, distinct collective spin excitations are still present. They show up in the transverse susceptibility as peaks in $\chi''(q\omega)$ with a maximum at $\omega=\omega_q$ and a half width $\Delta\omega_{1/2}$. For small $\omega<0.4t$, the peak width is twice the artificially introduced width $\Delta\omega_{1/2}=2\gamma$ (for $g=0.04t$). However, it increases sharply as ω is raised, beginning with $\omega\sim 0.4t$ ($|q|>0.7$ rad). Figure 2 shows the dispersion ω_q of the spin excitations for a system with $n=0.85$. The size of the vertical segments on the curve characterizes the peak width $\Delta\omega_{1/2}$ of the $\chi''(q\omega)$ curve for fixed q. The spin wave velocity $c=d\omega/dq|_{q=0}$ is somewhat higher in doped systems; this corresponds to the reduced dielectric gap Ud_0 with doping. The sudden broadening and asymmetry of the peak for large ω does not permit extending the ω_q curve into the region $q>1.2$ rad.

This picture is in qualitative agreement with measurements of $\chi''(q\omega)$ in LSCO at high frequencies, 25 meV $<\omega<200$ meV. [15] However, at low frequencies, $\omega<20$ meV, the susceptibility calculated for a doped antiferromagnetic state does not reproduce the low-energy peaks for the incommensurate quasimomenta $Q=(\pi\pm\delta,\pi)$ and $Q=(\pi,\pi\pm\delta)$ observed in LSCO.[8–10] In the two alternative approaches, these peaks are attributed either to different behavior of the Fermi boundary of the ''zero'' band for the different cuprates under antiferromagnetic correlation conditions,[8–10] or to an actual spin and charge superstructure, the so-called stripe phases observed in a number of cuprates.[38,40–42] An interpretation of the first type has one major defect. It is based on a zero susceptibility associated with the zero spectrum of ϵ_k. One of our purposes has been to study the effect of the Fermi surface on $\chi(k,\omega)$ when the band splits into upper and lower Hubbard subbands. It has been established previously[29–32] that under the conditions of two-dimensional antiferromagnetism, the Fermi surfaces are very sensitive to the parameter t', which accounts for diagonal jumps. When the sign of t' changes, the type of Fermi surface changes. For $t'>0$ doping forms hole pockets around $k\sim(\pm\pi/2,\pm\pi/2)$, and for $t'<0$, around $k\sim(\pm\pi,0)$, $(0,\pm\pi)$. Despite this difference, calculations of $\chi''(Q\omega)$ using Eqs. (23) and (25) for each of these systems, $t'>0$ or $t'<0$, give similar pictures of the spin excitations $\omega_q=cq$ with small variations in the velocity c, but do not reveal any peaks for incommensurate quasimomenta. The main influence on the variation in c is differences in the dielectric gap Δ of systems of different types and the reduction in Δ owing to doping.

The reason for this insensitivity of the dispersion of spin excitations to the details of the band energies when $\omega\ll t$ (contrary to the conclusions of Refs. 18–20) is that the collective spin mode develops from interband pairs, one particle form the upper band and the hole from the lower band. For them, the weighting functions (25) (the matrix elements of the transition) are of order ~ 1. When a dielectric gap ($E_{k+q,1}-E_{k,2}\geqslant 2\Delta\gg\omega$) is present, the details of the band have little influence on the interband contribution to Eq. (24) for $\chi^0(q,\omega)$. With doping and small q and ω, the interband contribution to $\chi^0(q,\omega)$ is still fundamental. In fact, for small ω, of all the particle–hole pairs $\{b^{\dagger}_{1k'\uparrow},b_{1k\downarrow}\}$ within a given lower band, only those pairs whose quasimomenta $k'=k+q$ and k lie within a narrow band near the Fermi boundary make a contribution to D_{ij}^0 [see Eq. (24)]. For light doping, these k and k' lie close to the nesting lines. But in this region, the matrix elements for one-band transitions and the corresponding weighting functions (25) are small, $R_{11}\ll|R_{12}|$. As a result of all this (small phase volume and weighting functions), in lightly doped systems with small $\omega\ll t$ and $|q|<1$, a collective mode with very little broadening is still attributable to interband pairs. Only for $\omega\geqslant 0.5t$ does the interaction with the single-band pairs cause a sharp broadening of the collective modes.

Therefore, the calculations do not yield incommensurate low-frequency peaks in $\chi''(q,\omega)$ within a picture of antiferromagnetically split bands for any type of Fermi surface. This contradicts the predictions of Refs. 18–20, which were based on cruder renormalizations of the zero susceptibility of the unsplit band ϵ_k.

4. SPIN SUSCEPTIBILITIES IN THE CASE OF SPIRAL STATES

In light of the above remarks, the explanation of the incommensurate peaks as a manifestation of nonuniform spin and charge structures observed in the cuprates[36–42] is plausible. Hartree–Fock model calculations[51–53] confirm the possibility of stabilizing these structures. More accurate calculations, in particular, ones that include valence-bond correlations, are desirable, but rather involved.

In this paper we limit ourselves to studying the susceptibility of the simplest spin structures—the spiral spin states of the Hubbard average field model. It is not clear whether these states have a direct connection with cuprates such as LSCO. As opposed to the oxides of nickel, in LSCO the incommensurate spin fluctuations may have a dynamic char-

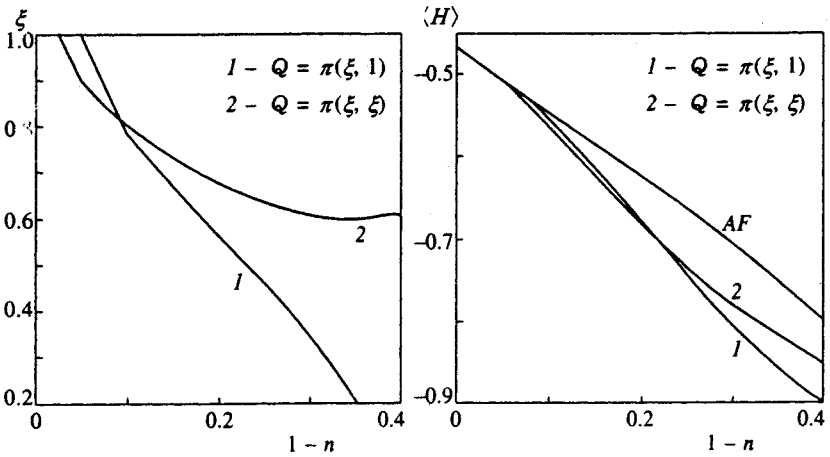

FIG. 5. The average energy $\langle H\rangle$ at a site (in units of t) and the parameter ξ characterizing the vector $Q=\pi(\xi,1)$ or $Q=\pi(\xi,\xi)$ of the spiral state with x- or xy symmetry, as functions of doping. The curve AF corresponds to the antiferromagnetic state of the average field with $Q=(\pi,\pi)$. The system parameters are: $U/t=8$, $t'/t=-0.05$.

acter without explicit signs of static stripe phases. The properties of an electronic system interacting with a spin system in a spiral state have been studied before.[54] Unlike the $s-d$ model examined there, with its two separate subsystems of spin and itinerant electrons, in the Hubbard model that we are studying the localized spins are formed directly from the electrons themselves. This changes the electronic spectrum and the spectrum of the magnetic fluctuations compared to Ref. 54.

Two types of spiral states of x- and xy-symmetry will be examined, with spirality vectors $Q=\pi(\eta,1)$ and $Q=\pi(\eta,\eta)$, respectively. The difference in the energies of the two types of spiral states of the average field is less than the expected gain in energy from the valence-bond correlations. Thus, it is impossible to establish a preference for one of the two types of states on the basis of the average energy alone. In the following some of the notation duplicates that of Sec. 2 but has a different meaning. For example, Q is the quasimomentum characterizing a spiral state, while q is the quasimomentum of spin excitation which changes within the limits of the entire Brillouin zone.

The spiral states with vector Q[44,55] are single-determinant states characterized by one-electron averages of the form

$$r_0=\langle c^\dagger_{n\sigma}c_{n\sigma}\rangle=\langle c^\dagger_{n\bar\sigma}c_{n\bar\sigma}\rangle,$$
$$\langle r_{\sigma\bar\sigma}\rangle=\langle c^\dagger_{n\sigma}c_{n\bar\sigma}\rangle=d_0e^{i\xi_\sigma Qn},\quad \xi_\sigma=\sigma/|\sigma|. \tag{39}$$

The linearized Hamiltonian of the Hubbard model for this class of states has the form

$$H_L=\sum_{k,\sigma}\epsilon_k c^\dagger_{k\sigma}c_{k\sigma}-Ud_0$$
$$\times\sum_{k,\sigma}c^\dagger_{k-Q/2,\sigma}c_{k+Q/2,\bar\sigma}+U(r_0^2-d_0^2). \tag{40}$$

Here $\bar\sigma=-\sigma$. The band energies $E_{k\lambda}$, $\lambda=1,2$ and the one-electron operators $b^\dagger{}_{k\lambda}$ of the Hamiltonian (40) are given by

$$E_{k1(2)}=\frac{1}{2}(\epsilon_{k-Q/2}+\epsilon_{k+Q/2})\mp g_k,$$
$$b^\dagger_{k\lambda}=\{c^\dagger_{k-Q/2,\uparrow},c^\dagger_{k+Q/2,\downarrow}\}_iU_{i\lambda}, \tag{41}$$

where

$$g_k=\sqrt{\delta\epsilon_k^2+U^2d_0^2},\quad \delta\epsilon_k=\frac{1}{2}(\epsilon_{k-Q/2}-\epsilon_{k+Q/2}), \tag{42}$$

and

$$U_{i\lambda}(k)=\begin{pmatrix}\cos\varphi_k & \sin\varphi_k\\ \sin\varphi_k & \cos\varphi_k\end{pmatrix},\quad \tan 2\varphi_k=Ud_0/\delta\epsilon_k. \tag{43}$$

Here k runs through all the values within the complete Brillouin zone. The definitions in Eqs. (41) and (42) are given in a form that is symmetric with respect to $\sigma\to-\sigma$.

Knowledge of the eigenstates and bands (41) and (42) makes it possible, in turn, to calculate the one-electron averages (39), in particular d_0:

$$d_0=\frac{1}{2N}\sum_k\frac{Ud_0(f_{1k}-f_{2k})}{2g_k}. \tag{44}$$

This closes the procedure for self-consistency, i.e., for minimizing the energy $\langle H\rangle$ with respect to variations in the functions for a fixed parameter Q. The subsequent minimization of $\langle H\rangle$ with respect to the spiral state parameter Q determines the optimum pitch of the spiral.

Figure 5 shows the parameter ξ which characterizes the spirality vector Q and the average energy for the two types of spiral states with $Q=\pi(\eta,1)$ and $Q=\pi(\eta,\eta)$ as functions of the doping.

When the linearized Hamiltonian is chosen in the form (40), the perturbation $V=H-H_L=H_U-(H_U)_L$ serves as the interaction which drives the collective spin excitations. The expression for H_U in terms of the fermi operators $b_{k\lambda}$ is derived directly and is given by

$$H_U=\frac{U}{N}\sum_{k_i,\lambda_i}U_{1\lambda1}(k_1)U_{2\lambda2}(k_2)U_{3\lambda3}(k_3)U_{4\lambda4}(k_4)$$
$$\times b^\dagger_{k_1\lambda_1}b_{k_2\lambda_2}b^\dagger_{k_3\lambda_3}b_{k_4\lambda_4}\delta(k_1-k_2+k_3-k_4). \tag{45}$$

The corresponding linearized operator $(H_U)_L$ is found in the usual way. Note that the operator $b^\dagger_{k\lambda}$ [see Eq. (41)] is not the eigenoperator of the quasimomentum associated with translation over the lattice period. Thus, the parameters k_i in

Eq. (45) are not true quasimomenta. Nevertheless, the δ-function has the same form as for an interaction which conserves the total quasimomentum (pseudo-quasimomentum) of the new particles.

For calculating the susceptibilities $\chi_{\alpha\beta}(q,\omega)=\langle\langle S^q_\alpha(\omega);S^{-q}_\beta\rangle\rangle$ in a system, whose ground state is spiral, we express the q-components of the spin and the densities in terms of the band operators $b_{k\lambda}$ for the spiral state:

$$S^q_\alpha=\frac{1}{2}\sum_{ij} r^q_{ij}(\sigma_\alpha)_{ij},\quad \rho=\frac{1}{2}\sum_{ij} r^q_{ij}(\sigma_0)_{ij},$$

$$r^q_{ij}=\frac{1}{N}\sum_k c^\dagger_{k+q,i}c_{kj}. \tag{46}$$

Here σ_α, $\alpha=1,2,3$, and σ_0 are the Pauli matrices. The indices $i,j=1,2$ have been introduced in place of $\sigma,\sigma'=\pm 1/2$. In the basis of the band Fermi operators (41), we have

$$r^q_{ij}=\frac{1}{N}\sum_{kk'} U_{i\lambda}(k)U_{j\lambda'}(k')b^\dagger_{k\lambda}b_{k'\lambda'}\delta \times(k-k'-q+(i-j)Q). \tag{47}$$

We introduce the following notation for the new operators that depend on the spatial harmonics

$$X^q_\alpha=\frac{1}{2}\sum_{ij}(\sigma_\alpha)_{ij}r^q_{ij},\quad \alpha=1,3,$$

$$X^q_1=\frac{1}{2}(r^{q-Q}_{12}+r^{q+Q}_{21}),\quad X^q_2=\frac{-i}{2}(r^{q-Q}_{12}-r^{q+Q}_{21}). \tag{48}$$

We shall characterize the operators defined this way by the common parameter q because of their uniform representation in terms of the spiral state band operators $b_{k\lambda}$,

$$X^q_\alpha=\frac{1}{2N}\sum_k\sum_{ij\lambda\lambda'}(\sigma_\alpha)_{ij}U_{i\lambda}(k+q)U_{j\lambda'}(k)b^\dagger_{k+q,\lambda}b_{k\lambda'}, \tag{49}$$

where $\alpha=0,1,2,3$. The representation (45) for H_U implies that, in the case of the ground spiral state, only those correlators of the operators X^q with the same q will be nonzero:

$$\langle\langle X^q_l(\omega);(X^{q'}_{l'})^\dagger\rangle\rangle=\delta_{qq'}G_{ll'}(q,\omega),\quad (X^q_\alpha)^\dagger=X^{-q}_\alpha. \tag{50}$$

This is yet another justification for introducing the operators (48).

We calculate these correlators using the same scheme as in Sec. 2. Given the representation (49) for X^q_l, we first find the equation for $\langle\langle\vartheta^q_{\mu\mu'}(kt);X^{-q}_{l'}\rangle\rangle$ and then for its ω-component, where $\vartheta^q_{\mu\mu'}(k)=b^\dagger_{k+q,\mu}b_{k\mu'}$. As a result of summing these $\langle\langle\vartheta^q_{\mu\mu'}(k\omega);X^{-q}_{l'}\rangle\rangle$ over k, μ, and μ' with the corresponding weighting functions from Eq. (49) we obtain the following system of algebraic equations for the $G_{ll'}(q\omega)$:

$$[\delta_{ll'}-UZ_{lm}\zeta_m]G_{ml'}=\frac{1}{2}Z_{ll'}, \tag{51}$$

where

$$\zeta_l=(1-2\delta_{l0})=\{-1,1,1,1\}_l,\quad l,m,l'=0,1,2,3,$$

$$Z_{ll'}(q)=-\frac{1}{2}\sum_{k\lambda\lambda'}\frac{f_{k+q,\lambda}-f_{k,\lambda'}}{E_{k+q,\lambda}-E_{k\lambda'}-\omega+i\gamma}F^l_{\lambda\lambda'}(k)\times(F^{l'}_{\lambda\lambda'}(k))^*, \tag{52}$$

and

$$F^l_{\lambda\lambda'}(k)=\sum_{ij}U_{i\lambda}(k+q)(\sigma_l)_{ij}U_{j\lambda'}. \tag{53}$$

Substituting Eq. (45) for $U_{i\lambda}$ into Eq. (53) yields compact matrix expressions for the matrices $F^l_{\lambda\lambda'}(k)$:

$$F^0=\sigma_0c_- -i\sigma_2s_-,\quad F^1=-\sigma_1c_+-\sigma_3s_+,$$
$$F^2=-\sigma_2c_-+i\sigma_0s_-,\quad F^3=\sigma_1s_++\sigma_3c_+, \tag{54}$$

where

$$c_\pm=\cos(\varphi_{k+q}\pm\varphi_k),\quad s_\pm=\sin(\varphi_{k+q}\pm\varphi_k).$$

Here the σ_l are the Pauli matrices and the φ_k are determined by Eq. (43). All the quantities in Eqs. (51)–(54) also depend implicitly on the ground state spirality parameter Q.

Calculating the $Z_{ll'}(q'\omega)$ using Eqs. (52)–(54) and solving the system of algebraic Eqs. (51), we find the correlators $G_{ll'}(q'\omega)$ for each of the three values of the argument $q'=q-Q,q,q+Q$. Recalling the definitions (47), (48), and (50), we obtain the following expressions for the unknown spin susceptibilities $\chi_{\alpha\beta}=\chi_{\alpha\beta}(q\omega)$ and the density correlations:

$$\chi_{zz}=\langle\langle S^q_z(\omega);S^{-q}_z\rangle\rangle=G_{33}(q\omega),$$
$$\langle\langle\rho^q_z(\omega);S^{-q}_z\rangle\rangle=G_{03}(q\omega),$$
$$\langle\langle S^q_z(\omega);\rho^{-q}_z\rangle\rangle=G_{30}(q\omega),$$
$$\langle\langle\rho^q_z(\omega);\rho^{-q}_z\rangle\rangle=G_{00}(q\omega), \tag{55}$$

$$\chi_{xx}=\chi_{yy}=\frac{1}{4}\{[G_{11}+G_{22}-i(G_{12}-G_{21})]|_{q+Q}+[G_{11}+G_{22}+i(G_{12}-G_{21})]|_{q-Q}\}, \tag{56}$$

and

$$\chi_{xy}=-\chi_{yx}=\frac{i}{4}\{[G_{11}+G_{22}-i(G_{12}-G_{21})]|_{q+Q}-[G_{11}+G_{22}+i(G_{12}-G_{21})]|_{q-Q}\}. \tag{57}$$

The arguments of the functions G_{ij} in each of the square brackets in Eqs. (56) and (57) are, respectively, $q+Q,\omega$ or $q-Q,\omega$. The indices x,y,z refer to the spin system of coordinates with an xy plane of rotation for the average spin of the spiral state.

If in calculating the $Z_{ll'}$ we include only the main contribution from the interband particle–hole pairs [$\lambda,\lambda'=1,2$ or 2,1 in Eq. (53)], then the determinant of the system goes to zero for $\omega\to 0$ and $\gamma\to 0$ because Eq. (44) is satisfied and the transverse components $G_{ij}(k\omega)$, $i,j=1,2$ will diverge for $k\to 0$ and $\omega\to 0$. According to the relations (56) and (57) between χ_{xx}, χ_{yy}, and $G(q\pm Q)$, this controls the possibility of the appearance of low-frequency peaks in $\chi''(q\omega)$ with incommensurate quasimomenta $q=\pm Q$. A complete calcu-

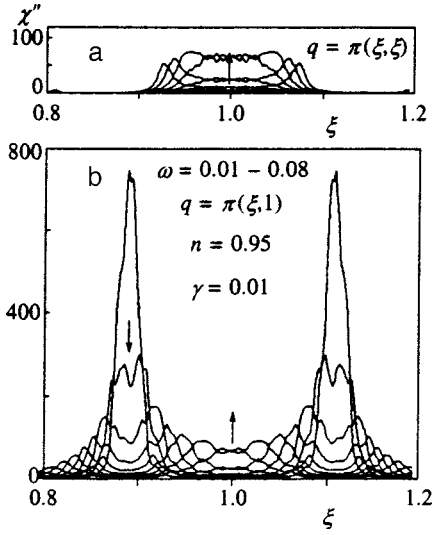

FIG. 6. $\chi''(q,\omega)$ as a function of the quasimomentum that varies along the diagonal, $q=\pi(\xi,\xi)$ (Fig. a), or along the x-axis, $q=\pi(\xi,1)$ (Fig. b) for a set of frequencies $\omega/t=0.01\rightarrow 0.09$. The calculation was done for a spiral state with doping $1-n=0.05$, corresponding to the vector $Q=\pi(0.892,1)$. The parameters are: $\gamma/t=0.01$, $U/t=8$, and $t'/t=-0.05$. χ'' is in units of $2t^{-1}$. For $\omega\rightarrow 0$ the peaks correspond to the incommensurate momenta $q_x=\pi(1\pm 0.108)$ on the x-axis.

lation according to Eqs. (55)–(57) and (51)–(53) confirms this.

Figure 6 shows the calculated $\chi''(q\omega)$ for $\omega=0.01\rightarrow 0.09$ at two cross sections of the phase plane, $q=\pi(\xi,1)$ and $q=\pi(\xi,\xi)$, for the spiral state with x-symmetry with a spirality vector $Q=\pi(0.892,1)$, which is optimal for $n=0.95$. The incommensurate peaks show up on the x axis at the points $q=Q$ and $[2(\pi,\pi)-Q]$ which are symmetric relative to the vector (π,π). Similarly, for the spiral state with the other symmetry with $Q=\pi(\eta,\eta)$, low-frequency peaks show up in $\chi''(q\omega)$ at incommensurate points which lie symmetrically on the diagonal relative to (π,π). The corresponding curves on the half interval are shown in Fig. 7. The narrow peak with $\xi=0.83$ is attributable to the contribution of intraband particle–hole pairs and depends on parallel segments of the Fermi surface of the lower Hubbard band for $n=0.95$. The type of Fermi surface has little effect on the main peaks for $q=\pm Q$.

As the frequency is increased, the main absorption peak splits. For a given ω, the maxima in the $\chi''(q,\omega)$ curve correspond to quasimomenta satisfying the condition $c^*|q-Q|=\omega$. The latter is analogous to the dispersion $\omega\propto c|q-(\pi,\pi)|$ of spin waves under antiferromagnetic spin ordering. For a system with $U/t\sim 8$ the velocity $c^*\sim 0.7t/a_0$ of the spin waves around the incommensurate vector Q is of the same order of magnitude as for the antiferromagnetic solution. Note that similar spin waves around an incommensurate Q have been observed in the nickel oxides,[36] where the existence of superstructures of inhomogeneous stripe phases has been demonstrated rigorously. Our results are not directly applicable to charge-inhomogeneous structures. Nevertheless, spiral states are interesting as the simplest charge-homogeneous systems that have a spin structure with an incommensurate period.

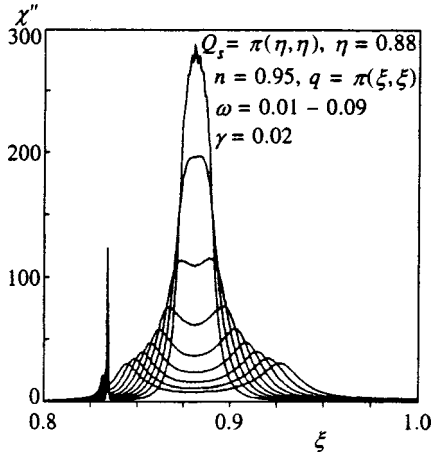

FIG. 7. As in Fig. 6, for $q=\pi(\xi,\xi)$ varying along the diagonal and spiral state of xy-symmetry with $Q=\pi(0.882,0.882)$. The region $\xi<1$ of only one of the two incommensurate peaks around (π,π) is shown. The position $q_{\max}(\omega)$ of the maxima of the $\chi''(q,\omega)$ curves for fixed ω corresponds to the spin-wave dependence $\omega=c|q_{\max}-Q|$.

As the doping is increased, the shift $\Delta Q=|Q-(\pi,\pi)|$ in the low-frequency peak of the susceptibility $\chi''(q,\omega)$ changes in accordance with the $Q(n)$ curves shown in Fig. 5a. Figure 8 shows the same $\chi''(q,\omega)$ curves as in Fig. 6, but for a system with $n=0.85$.

Yet another conclusion follows from a discussion of spiral states. The development of a spin structure along with the parameter t' that characterizes the diagonal jumps, in itself, changes the shape of the Fermi surface and this means that it can influence the low-energy properties of the system. As an illustration, Fig. 9 shows the Fermi surfaces and band energy levels $[E(k)-\mu]$ as functions of the two-dimensional parameter k, the pseudoquasimomentum of the one-electron eigenoperators $b_{k\lambda}$ of the spiral state. It is evident from Fig. 9 that the spiral spin ordering splits the van Hove singularities (VHS) of the one-electron spectrum in energy and introduces an asymmetry in their positions in

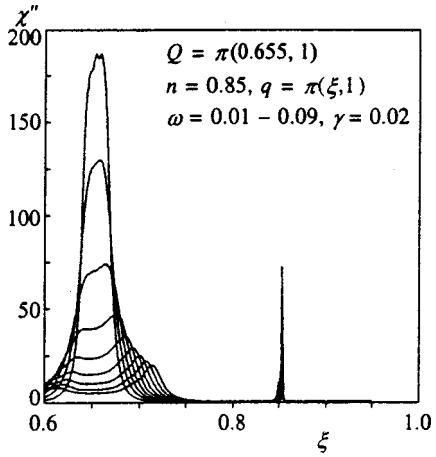

FIG. 8. As in Fig. 6, but for doping $1-n=0.15$, corresponding to the vector $Q=\pi(0.658,1)$ of the spiral state with the same x-symmetry. The region $\xi<1$ of only one of the two incommensurate peaks around (π,π) is shown. The peak at $\xi\sim 0.85$ is attributable to intraband particle–hole pairs and depends on the Fermi surface of the lower Hubbard band.

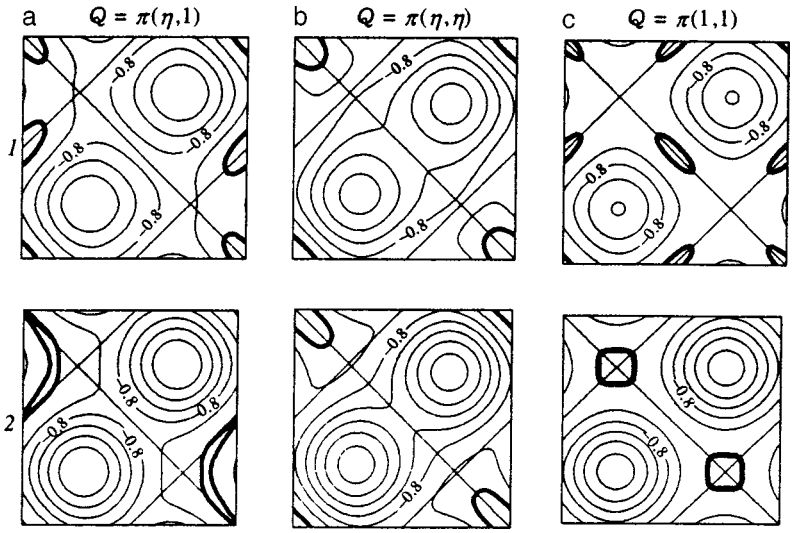

FIG. 9. Fermi surfaces (thick lines) and energy levels ($E_k - \mu$) of the lower Hubbard band for systems with $n=0.95$ in spiral states with x- and xy-symmetry or antiferromagnetic states of the average field in the (k_x, k_y) pseudoquasimomentum plane with $|k_{x(y)}| \leqslant \pi$. Figures a, b, and c correspond to $Q=\pi(\eta,1)$, $Q=\pi(\eta,\eta)$, and $Q=(\pi,\pi)$; *1* and *2* correspond to $t'/t=0.05$ or $t'/t=-0.05$. The diagonal straight lines correspond to nesting lines of the original band in these coordinates.

phase space. This behavior can be regarded as the electron analog of Jahn–Teller splitting of the VHS owing to lattice distortions.[2,56]

5. CONCLUSIONS

We have derived an expression for the spin susceptibility of the Hubbard model in a band approach with the RPA method using antiferromagnetically split Hubbard bands as a zeroth approximation. It has been shown that a collective mode of the spin excitations is formed by interband particle–hole pairs and, therefore, is insensitive (contrary to the predictions of the cruder RPA theory[18–20]) to the form and type of Fermi surface.

The calculation conveys well the characteristics of spin waves in a series of undoped and lightly doped cuprates. This serves as an additional argument that the band approach in a picture of split upper and lower Hubbard subbands can adequately describe the low-energy spin dynamics of a correlated system. This approach should be supplemented by including valence-band correlations, which induce[30] superconducting pairing, and possible charge and spin inhomogeneities for describing the superstructures that exist in a number of cuprates.

We have demonstrated the impossibility of explaining the incommensurate inelastic neutron scattering peaks in LSCO in terms of the properties of the Fermi surface of a homogeneous antiferromagnetic state of the average field. The earlier conclusion[36–42] that the source of the spin fluctuations with an incommensurate momentum is inhomogeneous structures has been confirmed by a model calculation of the susceptibility for the simplest structures with an incommensurate spin-ordering—charge-homogeneous spiral states.

This work was supported by the Russian Fund for Fundamental Research (Projects No. 7-03-33727A and No. 96-15-97492), as well as by the International Scientific-Technical Center (Grant No. 872). We thank R. O. Zaĭtsev and V. Ya. Krivnov for useful comments and discussions.

APPENDIX

In the self-consistent solution of the problem with the zeroth-order linearized Hamiltonian (9), the perturbation is an operator with four- and two-fermion contributions. In the basis of the eigenoperators (10), they have the form

$$V=V_4+V_2, \quad V_4=H_U, \quad V_2=-(H_U)_L, \tag{A1}$$

with

$$V_4=\frac{U}{N}\sum_{m=1,2}\sum_{kk'q}{}^{F} M^m_{\lambda\lambda'}(qk\sigma\bar{\sigma})M^m_{\nu\nu'}(qk'\sigma\bar{\sigma}) \times b^\dagger_{\underline{k+q}\lambda\sigma}b_{\underline{k'+q}\nu\sigma}b^\dagger_{k'\nu'\bar{\sigma}}b_{k\lambda'\bar{\sigma}}, \tag{A2}$$

and

$$V_2=-Ur_0N+Ud_0 \times \sum_{\sigma\lambda\lambda'}\xi_\sigma\sum_k{}^{F} M^m_{\lambda\lambda'}(qk\sigma\sigma)b^\dagger_{k\lambda\sigma}b_{k\lambda'\sigma}. \tag{A3}$$

Here $2r_0$ and d_0 are the average and alternating spin densities of the particles at a node; $\xi_\sigma=\sigma/|\sigma|$; and, $\underline{k+q}=k+q+(\pi,\pi)i_{kq}\in F$ is the quasimomentum, normalized to the magnetic Brillouin zone with the corresponding normalization index i_{kq}. The functions $M^m_{\lambda\lambda'}(qk\sigma\sigma')$ are given by Eq. (17). In particular, for $\sigma'=\bar{\sigma}=-\sigma$ we obtain

$$M^m_{\lambda\lambda'}(qk\sigma\bar{\sigma})=\begin{pmatrix} c_+ & \xi_\sigma s_+ \\ -\xi_\sigma s_+ & c_+ \end{pmatrix}_{\lambda\lambda'}\delta^{(2)}(m+i_{kq}) + \begin{pmatrix} \xi_\sigma s_- & c_- \\ c_- & -\xi_\sigma s_- \end{pmatrix}_{\lambda\lambda'}\delta^{(2)}(m+i_{kq}+1), \tag{A4}$$

where

$$c_\pm=\cos(\varphi_{\underline{k+q}}\pm\varphi_k), \quad s_\pm=\sin(\varphi_{\underline{k+q}}\pm\varphi_k). \tag{A5}$$

Here $\delta^{(2)}$ is the δ-function *modulo* 2. The averages r_0 and d_0 are determined by standard methods.

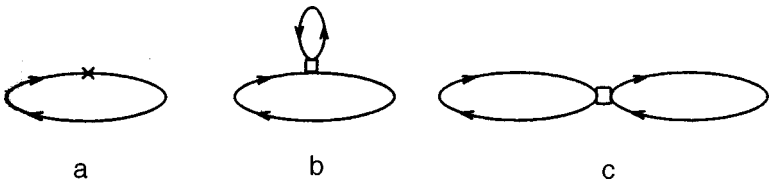

FIG. 10. First-order diagrams in U for the transverse polarization operator.

In the calculation of the commutator (22), the contribution $[\vartheta, V_2]$ cancels with the corresponding terms in $[\vartheta, V_2]$ as $n_{k\lambda\sigma} \to f_{k\lambda\sigma}$. The result can be explained in terms of the language of diagrams in first order in U for the transverse polarization operator.[44] Thus, in the diagrams of Fig. 10, the contributions from a and b cancel out and only the contribution from the ladder class c remains. The cross and square in Fig. 10 correspond to interactions of V_2 and V_4, respectively. Summing the ladder diagrams of all orders in the basis of the zeroth-order Hamiltonian H_L and the perturbation V gives a result identical to Eq. (23).

It has been shown previously[21–24] that this class of diagrams is insufficient if the zeroth-order basis $\{c_{k\sigma}\}$ of the initial band ϵ_k is used with nesting and van Hove singularities from the four points A, B, C, and D: $k=(\pm\pi,0)$, $(0,\pm\pi)$. However, in the basis (10) of split subbands, the zero vertex parts for the points A, B, C, and D and all the nesting lines goes to zero for the dangerous diagrams. The latter include the diagrams with propagators $G(k\lambda\omega)G(k'\lambda'\omega')$ corresponding to a single subband $\lambda=\lambda'$. (For the interband contributions $\lambda\neq\lambda'$ the resulting energy denominator has no singularities owing to the Hubbard gap $E_{k2}-E_{k1}>2Ud_0$.) In fact, for $q=0$ or (π,π) and $k,k'=\underline{k+q}$ lying on the nesting lines $|k_x \pm k_y|=\pi$, we have

$$\varphi_k=\varphi_{k'}=-\pi/4, \quad c_+=s_-=0, \quad c_-=-s+=1, \quad \text{(A6)}$$

which, with Eq. (61), gives $M^m_{\lambda\lambda}(qk\sigma\underline{\sigma})=0$ for these k and q. In this regard, we can limit ourselves to summing the ladder diagrams with a zeroth approximation for the vertex part, which corresponds to the result (23) for the transverse susceptibility. This is not surprising, since a new basis with disrupted symmetry was chosen precisely to eliminate the instability with respect to the spin density waves observed in the first expansions with the original band.[21–24]

The calculations of the susceptibility for the longitudinal components of the spin and density similar to the derivation of Eq. (23). Of these,

$$B_{ij}(q\omega)=\begin{pmatrix} \langle\langle S_z(q\omega);S_z(q)\rangle\rangle & \langle\langle S_z(q\omega);\rho(\tilde{q})\rangle\rangle \\ \langle\langle \rho(\tilde{q}\omega);S_z(q)\rangle\rangle & \langle\langle \rho(\tilde{q}\omega);\rho(\tilde{q})\rangle\rangle \end{pmatrix}_{ij},$$

$$q\in F, \quad i,j=1,2, \quad \text{(A7)}$$

determined within the magnetic Brillouin zone and analogous to the correlators $B_{ij}(\tilde{q}\omega)$ for $\tilde{q}=q+(\pi,\pi)\notin F$, are nonzero. Repeating the derivation as for Eqs. (23), we obtain the following expressions for the B_{ij}:

$$B_{ij}(q,\omega)=\frac{1}{2}\{[I-UB^0\sigma_3]^{-1}B^0\}_{ij} \quad \text{(A8)}$$

and

$$B_{ij}(\tilde{q},\omega)=\frac{1}{2}\{[I-U\sigma_1B^0\sigma_1\sigma_3]^{-1}\sigma_1B^0\sigma_1\}_{ij}. \quad \text{(A9)}$$

Here σ_1 and σ_3 are the Pauli matrices and the two-dimensional matrix B^0_{ij} is given by

$$B^0_{ij}=-\frac{1}{N}\sum_{\lambda\lambda'}\sum_k{}^F \frac{f_{k+q,\lambda}-f_{k\lambda'}}{E_{k+q,\lambda}-E_{k\lambda'}-\hbar\omega+i\gamma}P^{ij}_{\lambda\lambda'},$$

where

$$P^{11}_{\lambda\lambda'}=\begin{pmatrix} c_-^2 & s_-^2 \\ s_-^2 & c_-^2 \end{pmatrix}_{\lambda\lambda'}, \quad P^{22}_{\lambda\lambda'}=\begin{pmatrix} s_+^2 & c_+^2 \\ c_+^2 & s_+^2 \end{pmatrix}_{\lambda\lambda'},$$

$$P^{12}_{\lambda\lambda'}=P^{21}_{\lambda\lambda'}=\begin{pmatrix} c_-s_+ & s_-c_+ \\ -s_-c_+ & -c_-s_+ \end{pmatrix}_{\lambda\lambda'},$$

$$c_\pm=\cos\varphi_\pm, \quad s_\pm=\sin\varphi_\pm, \quad \varphi_\pm=\varphi_{k+q}\pm\varphi_k.$$

$c_\pm$, $s_\pm$, and φ_k are determined by Eqs. (62) and (11); $E_{k\lambda}$ and $f_{k\lambda}$ $(\lambda=1,2)$ are the band energies and Fermi functions of the upper and lower Hubbard bands. It is important that the determinants of the matrices in parentheses in Eqs. (65) and (66) do not go to zero as $\omega\to 0$. This means that there are no low-frequency collective fluctuation modes of the longitudinal spin and density components.

*)E-mail: movchin@center.chph.ras.ru

[1] E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
[2] R. S. Markievicz, J. Phys. Chem. Solids **58**, 1179 (1997).
[3] Z.-X. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995).
[4] R. J. Birgeneau and G. Shirane, in *Physical Proprties of High Temperature Superconductors. I*, ed. by D. M. Ginzberg, World Scientific, Singapore (1989).
[5] D. S. Marshall, D. S. Dessau, A. G. Loeser *et al.*, Phys. Rev. Lett. **76**, 4841 (1996).
[6] A. G. Loeser, Z.-X. Shen, D. S. Dessau *et al.*, Science **273**, 325 (1997).
[7] H. Ding, T. Yokoya, J. C. Campuzano *et al.*, Nature (London) **382**, 51 (1996).
[8] T. E. Mason, G. A. Aepli, S. M. Hayden *et al.*, Phys. Rev. Lett. **71**, 919 (1993).
[9] S. W. Cheong, G. Aeppli, T. E. Mason *et al.*, Phys. Rev. Lett. **67**, 1791 (1991).
[10] K. Yamada, S. Wakimoto, G. Shirane *et al.*, Phys. Rev. Lett. **75**, 1526 (1995).
[11] J. Rossat-Mignod, L. P. Renault, C. Vettrier *et al.*, Physica C **185**, 86 (1991).
[12] L. P. Regnaught, P. Bourges, P. Burlet *et al.*, Physica B **213**, 48 (1995).
[13] P. Bourges, H. Casalta, A. S. Ivanov *et al.*, Phys. Rev. Lett. **79**, 4906 (1997).
[14] S. M. Hayden, G. Aeppli, H. A. Mook *et al.*, Phys. Rev. Lett. **76**, 1344 (1997).
[15] P. Bourges, H. F. Fong, L. P. Regnault *et al.*, Phys. Rev. B **56**, R11436 (1997).

[16] H. F. Fong, B. Keimer, D. L. Milius *et al.*, Phys. Rev. B **54**, 6708 (1996); Phys. Rev. Lett. **78**, 713 (1997).
[17] P. Dai, H. Mook, and F. Dogan, Phys. Rev. Lett. **80**, 1738 (1998).
[18] D. Z. Liu, Y. Zha, and K. Levin, Phys. Rev. Lett. **75**, 4310 (1995).
[19] Q. Si, Y. Zha, K. Levin *et al.*, Phys. Rev. B **47**, 9055 (1993).
[20] P. Benard, L. Chen, and A.-M. S. Tremblay, Phys. Rev. B **47**, 587 (1993).
[21] I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **93**, 1487 (1987) [Sov. Phys. JETP **66**, 848 (1987)].
[22] I. E. Dzyaloshinskiĭ and V. M. Yakovenko, Zh. Éksp. Teor. Fiz. **94** (4), 344 (1988) [Sov. Phys. JETP **67**, 844 (1988)].
[23] I. E. Dzyaloshinskiĭ and E. I. Kats, Zh. Éksp. Teor. Fiz. **62**, 1105 (1972) [Sov. Phys. JETP **35**, 584 (1972)].
[24] H. Shulz, Europhys. Lett. **4**, 609 (1987).
[25] J. Ruvalds and A. Virosztek, Phys. Rev. **4**, 609 (1991); **42**, 4064 (1990).
[26] E. Dagotto, A. Nazarenko, and A. Moreo, Phys. Rev. Lett. **74**, 310 (1995).
[27] A. Nazarenko and E. Dagotto, Phys. Rev. Lett. **74**, 310 (1995).
[28] R. S. Markiewicz, Phys. Rev. B **56**, 9091 (1997).
[29] A. A. Ovchinnikov and M. Ya. Ovchinnikova, Phys. Lett. A **249**, 531 (1998).
[30] A. A. Ovchinnikov, M. Ya. Ovchinnikova, and E. A. Plekhanov, JETP Lett. **67**, 369 (1998); Zh. Éksp. Teor. Fiz. **114**, 985 (1998) [JETP **87**, 534 (1998)]; Zh. Éksp. Teor. Fiz. **115**, 649 (1999) [JETP **88**, 356 (1999)].
[31] N. M. Plakida, V. S. Oudovenko, R. Horsch *et al.*, Phys. Rev. B **55**, 11 997 (1997).
[32] R. O. Kuzin, R. Hayn, A. F. Barabanov *et al.*, Phys. Rev. B **58**, 6194 (1998).
[33] I. A. Misurkin, A. A. Ovchinnikov, and G. A. Vinogradov, Int. J. Quantum Chem. **9**, 605 (1975).
[34] D. Pines, Z. Phys. B **103**, 129 (1997).
[35] J. Hubbard and K. P. Jain, J. Phys. C **1**, ser. 2, 1650 (1968).
[36] J. M. Tranquada, P. Wocher, and D. J. Buttrey, Phys. Rev. Lett. **79**, 2133 (1997); Phys. Rev. B **55**, R6113 (1997).
[37] A. Bianconi and M. Missori, in *Phase Separation in Cuprate Superconductors*, ed. by E. Sigmound, K. A. Muller; Springer-Verlag, Berlin (1994).
[38] J. M. Tranquada, B. J. Sternlieb, J. D. Axe *et al.*, Nature (London) **375**, 561 (1995).
[39] B. Buchner, M. Brener, and A. Freimuth, Phys. Rev. Lett. **73**, 1841 (1994).
[40] J. M. Tranquada, D. J. Axe, M. Ichikawa *et al.*, Phys. Rev. B **54**, 7489 (1996); Phys. Rev. Lett. **78**, 338 (1997).
[41] X. L. Dong, Z. F. Dong, B. R. Zhao *et al.*, Phys. Rev. Lett. **80**, 2701 (1998).
[42] A. Bianconi, N. L. Saini, A. Lanzara *et al.*, Phys. Rev. Lett. **76**, 4312 (1996).
[43] T. Izuyama, T. Kim, and R. Kubo, J. Phys. Soc. Jpn. **18**, 1025 (1963).
[44] Yu. A. Izyumov, M. I. Katsnel'son, and Yu. N. Skryabin, *The Magnetism of Collective Electrons* [in Russian], Nauka, Moscow (1994).
[45] A. A. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **110**, 342 (1996) [JETP **83**, 184 (1996)]; Zh. Éksp. Teor. Fiz. **112**, 1409 (1997) [JETP **85**, 707 (1997)].
[46] U. Trapper, D. Ihle, and H. Fenke, Phys. Rev. B **52**, R11 553 (1995).
[47] G. Baumgartel, J. Schmalian, and K. H. Benemann, Europhys. Lett. **24**, 601 (1993).
[48] I. W. Sumarlin, J. W. Lynn, T. Chattapadhyay *et al.*, Phys. Rev. B **51**, 5824 (1995).
[49] H. B. Schuttler and A. J. Fedro, Phys. Rev. B **45**, 7588 (1992).
[50] J. H. Jefferson, H. Eskes, and L. F. Feiner, Phys. Rev. B **45**, 7959 (1992).
[51] T. Mizokawa and A. Fujimori, Phys. Rev. **56**, 11920 (1997).
[52] J. Zaanen and M. L. Horbach, Phys. Rev. **53**, 8671 (1996).
[53] V. J. Emery and S. A. Kivelson, Physica C **209**, 597 (1993).
[54] Yu. A. Izyumov and V. M. Laptev, Int. J. Mod. Phys. B **4**, 447 (1990).
[55] F. Hu, S. K. Sarker, and C. Jayaprakash, Phys. Rev. B **50**, 17901 (1994).
[56] R. S. Markiewicz, Physica C **255**, 211 (1995).

Translated by D. H. McNeill

Thermodynamics of a vortex system in a thin superconducting film with radiation defects

A. N. Artemov*)

Donetsk Physicotechnical Institute, National Academy of Sciences of Ukraine, 340114 Donetsk, Ukraine
(Submitted 8 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1081–1090 (September 1999)

The effect of radiation defects on the thermodynamics of a system of Pearl vortices in a thin superconducting film is examined. The scenario for a Kosterlitz–Thouless transition in this system is shown to depend on the defect concentration n_d. At low concentrations, the transition takes place continuously, while at high concentrations, a range of temperatures exists in which there are two metastable states. The concentrations of free vortices and of vortices captured by defects are calculated as functions of temperature for different defect concentrations n_d. A phase diagram is constructed for the vortex system in the $n_d - T$ plane. © *1999 American Institute of Physics.* [S1063-7761(99)02209-X]

1. INTRODUCTION

In recent years there has been persistent interest in research on layered superconductors with radiation (columnar) defects.[1–5] These defects are regions with sizes on the order of the coherence length ξ of the superconductor, within which superconductivity is completely suppressed. This interest originates in the possibility of controlling the properties of the superconductor by introducing this type of defects, which are efficient pinning centers. It has been shown[1–3] that the critical superconducting current can be raised significantly by this method.

On the other hand, layered superconductors are systems subject to strong influence by thermal fluctuations. They show up, for example, in the existence of the so-called line of irreversibility in layered superconductors,[6] which is also associated with the phenomenon of pinning. This is a line in the magnetic field–temperature plane, for which the magnetization process in the superconductor becomes reversible when it is crossed. We believe that the reversible behavior of the magnetic flux in superconductors with pinning is related to a Kosterlitz–Thouless (KT) transition[7,8] in a system of magnetic vortices created by this flux.

KT transitions take place in two-dimensional systems in which topological defects with a Coulomb interaction can exist. An example of a model two-dimensional superconducting system is layered superconductors without Josephson coupling between layers with two-dimensional magnetic vortices as topological defects. A KT transition in such a system is caused by two effects: an instability of the vortex dipoles against dissociation in the gas of free vortices which develops in the system above a temperature[9,10]

$$T_{\mathrm{KT}}=\frac{\phi_0^2}{16\pi^2\Lambda(T_{\mathrm{KT}})}, \tag{1}$$

and collective effects in the system of free vortices. Here ϕ_0 is the quantum of magnetic flux, $\Lambda=2\lambda^2/s$, λ is the London length, and s is the period of the layered system.

A thin superconducting film of thickness $d \ll \lambda$ is not a strictly two-dimensional system. Vortices in these systems were first examined by Pearl.[11] The logarithmic interaction of the vortices in a film is bounded by the large but finite effective Pearl length $\Lambda=2\lambda^2/d$. Nevertheless, it has been shown[12] that processes can take place in a system of Pearl vortices which ensure that the system behaves in a manner similar to a KT transition. These are the same instability and collective effects which cause a KT transition in a two-dimensional system. In a system of Pearl vortices, however, the correlation length cannot exceed Λ, while in a two-dimensional system it approaches infinity as the temperature T_{KT} is approached from above. Thus, the phenomena in a Pearl film which are referred to as a KT transition are not a phase transition in the strict sense.

Since the thermodynamic behavior of systems of two-dimensional and Pearl vortices is determined by the same processes, it is natural to expect that radiation defects will also have the same effect on this behavior. Their role is to capture and confine Pearl vortices, limiting their mobility. The absence of a normal core in a vortex captured by a radiation defect makes this state more favorable energetically, and this has a fundamental effect on the course of the processes that create a KT transition in a defective superconductor.

In this paper we examine the effect of radiation defects on the KT transition in a thin superconducting film. It is shown that, depending on the defect concentration n_d, three different scenarios for the phase transition can occur. For low n_d, below a certain critical concentration n_{d1}, the transition takes place continuously as a second-order transition. At intermediate densities, $n_{d1}<n_d<n_{d2}$, there are two thermodynamically equilibrium states of the free vortices, individual and collective. Here the phase transition takes place as a first-order transition and hysteresis should be observed in the temperature dependence of the resistance. At high defect concentrations, $n_d>n_{d2}$, the lower stability boundary of the metastable states is shifted almost to $T=0$.

2. FREE ENERGY OF THE VORTEX SYSTEM

We consider Pearl vortices as classical massless particles. They can be in a free state or be captured by radiation defects. In order to describe the partition function of a thermodynamic system of this type, one of two equivalent approaches can be used.

The free and trapped vortices can be treated as two subsystems in thermal and chemical contact. The free vortices in such a system can appear and disappear through the dissociation and recombination of vortical dipoles, and also shift into or leave the subsystem of vortices that have been trapped by defects. In equilibrium, the temperatures and chemical potentials of the subsystems are equal.

The other approach is to consider the processes taking place in the vortex system as ''chemical reactions.'' In such a system, the annihilation of two oppositely oriented free vortices or of a free and a trapped vortex can occur, as can the capture of a free vortex by an empty defect, as well as ''reactions'' in the opposite direction. Here a restriction is imposed on the system whereby the sum of the empty defects, N_{d0}, and of the defects that have captured a flux quantum, $N_{t+}+N_{t-}$, equals the total number N_d of defects in the system.

We write the partition function of the vortex system in a film with defects by analogy with a chemical system. In a real film, the defects form a random configuration specified by the set of their coordinates $\{\mathbf{R}_\alpha\}$. Let there be N_+ and N_- free vortices with the two orientations, N_{t+} and N_{t-} vortices captured by defects, and $N_d > N_{t+}+N_{t-}$ radiation defects. The number of vortices must be subject to the condition $N_++N_{t+}=N_-+N_{t-}$, which follows from the conservation law for topological charge. We shall not take this into account explicitly, since the symmetry of the equilibrium state of the system, which is all we are interested in, imposes a more severe restriction: $N_+=N_-$ and $N_{t+}=N_{t-}$. Then the partition function is given by

$$
\begin{aligned}
&Z(N_\pm, N_{t\pm}) \\
&\quad = \frac{1}{N_+!N_-!N_{t+}!N_{t-}!}\,\frac{1}{(N_d-N_{t+}-N_{t-})!} \\
&\qquad \times \exp\{-\beta(N_++N_-)E_0\}\,\mathrm{Tr}\exp\Bigg\{-\sum_{ij}\,\beta U(\mathbf{x}_i-\mathbf{x}_j) \\
&\qquad -\sum_{i\alpha}\,\beta U(\mathbf{x}_i-\mathbf{R}_\alpha)-\sum_{\alpha\gamma}\,\beta U(\mathbf{R}_\alpha-\mathbf{R}_\gamma)\Bigg\}. \qquad (2)
\end{aligned}
$$

Here the symbol ''Tr'' denotes a sum over all possible states of the vortices in the system:

$$
\mathrm{Tr}\equiv\left(\int\frac{d\mathbf{x}}{\pi\xi^2}\right)^{N_++N_-}\left(\sum_{R_\alpha}\right)^{N_{t+}+N_{t-}},
$$

where $\beta=1/T$, E_0 is the energy of the vortex core, $\pi\xi^2$ is the size of the spatial cell occupied by a single vortex, and U is the interaction energy of free vortices located at the points $\mathbf{x}_i$ with vortices captured by defects lying at the points $\mathbf{R}_\alpha$. The second cofactor in Eq. (2) is related to the identical nature of empty defects, which we treat as one of the components of the chemical system and is simply $1/(N_{d0}!)$, including the restriction mentioned above. We have neglected the contribution to the partition function from vortex dipoles associated with the polarization of the medium and the interaction of vortices with empty defects, which will be discussed below.

The free energy of the system depends on the coordinates of the defects and should be averaged over their positions. Only the configuration part of the energy needs to be averaged, since the entropy part is independent of the coordinates of the defects. To calculate the configuration energy, we formally expand the exponent in Eq. (2) in a series. The logarithm of the partition function is expanded in a series of connected diagrams.[13] For a gas of free vortices, the integral is taken over the positions of all the vertices in the diagrams. If the film contains defects at the points R_α, then diagrams accounting for the interaction with the vortices trapped in the defects, whose coordinates are fixed, show up in the sequence. The energy of these configurations depends on the coordinate of the captured vortices and cannot be calculated in general. In order to take the average of the configuration energy, we propose that all the defects be distributed with uniform probability over the entire plane of the sample, independently of one another. We take the average by integrating the series obtained formally above with respect to the coordinates of the trapped vortices and dividing each integral by the area S. As a result of this operation, the contribution of the trapped vortices, which were attached to definite points R_α in a specific configuration, is formally included in the free energy of the system on an equal footing with the contribution from the free vortices. The only difference is that the state of the vortices trapped by the defects is more energetically favorable because of the zero energy of the core.

Now it is easy to calculate the free energy of the system of vortices. Since we are mainly interested in collective effects in the vortex system, we restrict ourselves to summing the sequence of ring diagrams.[13]

The characteristic feature of systems in which collective effects predominate is that the integral $J=\int d\mathbf{r}U(r)$ corresponding to the simplest diagram diverges.[14] This means that even for a low concentration of vortices it is impossible to limit ourselves to their interaction with a particular number of nearest neighbors, but the interactions of each with all the others have to be taken into account. The ring diagrams are the principal sequence in the expansion of the configuration energy in this case.

The situation is different when a vortex interacts with an empty defect. The interaction energy of a Pearl vortex with an empty cylindrical defect has been calculated elsewhere.[15] It falls off with distance from a defect much more rapidly than the interaction energy with a vortex. In this case, the integral J converges and the configuration energy can be expanded in a series with respect to the concentration of vortices and empty defects. In the equilibrium equation it yields terms proportional to the concentrations, which are small compared to the logarithms of the concentrations and can be omitted. Thus, in order to simplify the formulas, from the beginning we neglect the contribution of the interaction

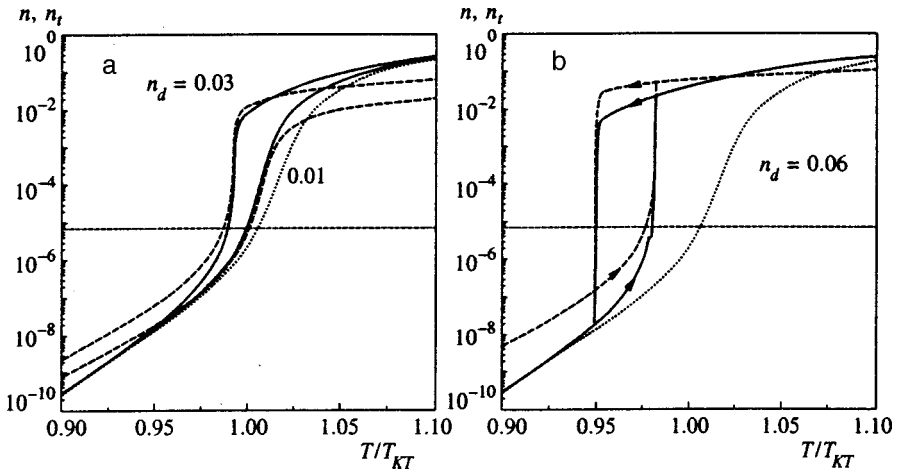

FIG. 1. The concentrations of free (smooth curves) and defect-trapped (dashed curves) vortices as functions of temperature in the case of a continuous transition a and the case of a first-order transition b. The thin dashed horizontal line indicates the total vortex concentration in the system for which the Debye shielding length δ equals the Pearl length Λ.

between the vortices and empty defects to the partition function.

The free energy of a system of Pearl vortices has been calculated in the ring approximation.[14] We need only replace the vortex concentration in the configuration part by the sum of the concentrations of the free vortices and the vortices trapped in defects. Introducing dimensionless notation for the concentrations of the vortices, $n=(N/S)\pi\xi^2$, and the free energy density, $f=(F/S)\pi\xi^2$, we write the free energy of the system in the form

$$f=n_+(\ln n_+-1)+n_-(\ln n_--1)+n_{t+}(\ln n_{t+}-1)+n_{t-}\times(\ln n_{t-}-1)+(n_d-n_{t+}-n_{t-})(\ln(n_d-n_{t+}-n_{t-})-1)+p(n_++n_-+n_{t+}+n_{t-})(1-\ln 4p(n_++n_-+n_{t+}+n_{t-}))+\frac{1}{4\Lambda^2}G[16p\Lambda^2(n_++n_-+n_{t+}+n_{t-})]+(n_++n_-)pe_0, \quad (3)$$

where

$$G[x]=\frac{1}{2}\ln\frac{x}{4}+\sqrt{|1-x|}\begin{cases}\arctan\dfrac{1}{\sqrt{x-1}}-\dfrac{\pi}{2}, & x\geqslant 1,\\ \dfrac{1}{2}\ln\dfrac{1+\sqrt{1-x}}{1-\sqrt{1-x}}, & x\leqslant 1.\end{cases}$$

Here $pe_0=\beta E_0$ and $p=\phi_0^2/16\pi^2\Lambda T$.

3. EQUILIBRIUM STATE OF THE VORTEX SYSTEM

A vortex system in a thin film is a system with a variable number of particles. The equilibrium number of particles in such a system must be determined from the condition of a minimum free energy, while the equilibrium chemical potential vanishes. Thus, we obtain the equilibrium condition for the system by equating to zero the derivatives of the free energy with respect to the concentration of free vortices, $n_\pm$, and the concentration of vortices trapped by defects, $n_{t\pm}$:

$$\ln n_\pm - p\ln 4p(n_++n_-+n_{t+}+n_{t-})+4pG'\times[16p\Lambda(n_++n_-+n_{t+}+n_{t-})]+pe_0=0,$$

$$\ln n_{t\pm}-\ln(n_d-n_{t+}-n_{t-})-p\ln 4p(n_++n_-+n_{t+}+n_{t-})+4pG'[16p\Lambda(n_++n_-+n_{t+}+n_{t-})]=0. \quad (4)$$

Subtracting the equation for n_- from that for n_+, we obtain $n_+=n_-$ in the equilibrium state. The two other equations give $n_{t+}=n_{t-}$. This reduces the number of equilibrium equations to two. In the following we shall omit the ''+'' and ''−'' subscripts.

Subtracting the equation for n_t from that for n, we obtain a relationship between the concentrations of free and trapped vortices,

$$n_t=\frac{nn_d}{2n+e^{-pe_0}}.$$

This equation shows that $n_t\to 0$ if the concentration n of free vortices approaches zero more rapidly than the exponent in the denominator. If, on the other hand, the exponent is substantially smaller than n, then the vortices tend to fill all the defects. A study of the equilibrium behavior of the vortex system has shown that the first of these asymptotic properties of the subsystem of trapped vortices shows up in the individual vortices and the second in the collective state of the vortex system.

The solutions of this system of equations for different defect concentrations n_d are plotted in Figs. 1 and 2. The temperature variations in the concentrations of the free (smooth curve) and trapped (dashed curve) vortices at relatively low defect concentrations $n_d<n_{d1}$ are discontinuous [Fig. 1(a)]. A sharp change in the concentrations near T_{KT} is associated with the development of an instability similar to that predicted by Kosterlitz and Thouless[8] for strictly two-dimensional systems. The same sort of instability is observed in perfect films (dotted curve), but when defects are present it is shifted toward low temperatures. It is caused by a transition of the vortex system into a collective state characterized by a Debye shielding length $\delta=\xi/\sqrt{8p(n+n_t)}$ for the vortex interaction that has become shorter than the effective Pearl shielding length Λ and the interaction energy of two vortices that depends on the concentration of vortices in the

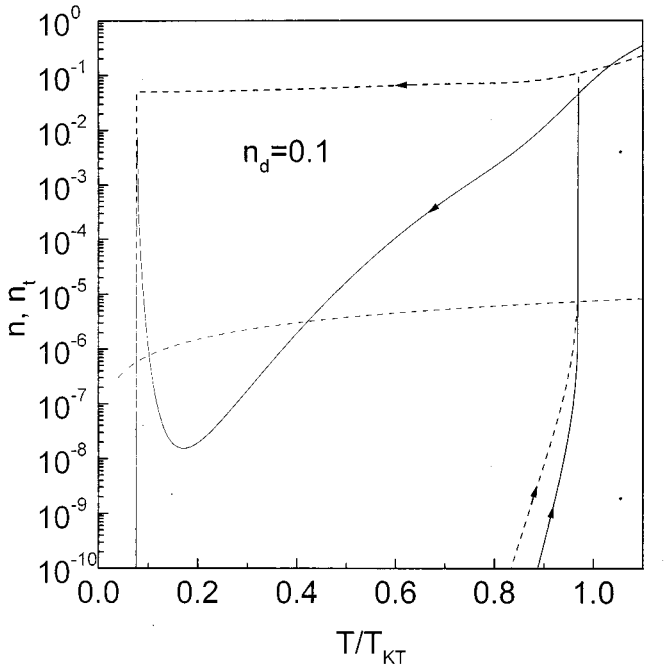

FIG. 2. The concentrations of free (smooth curves) and defect-trapped (dashed curves) vortices as functions of temperature in the case of a high defect concentration in the system, i.e., $n_d > n_{d2}$. The thin dashed line indicates the total vortex concentration in the system for which the Debye shielding length δ equals the Pearl length Λ.

system. Here additional creation of new free vortices becomes favorable, since the interaction energy of the vortices decreases as their concentration rises.

In Figs. 1 and 2 the thin dashed lines represent the total concentration of free and trapped vortices for which $\delta = \Lambda$. These lines arbitrarily separate the domains of the individual (below) and collective (above) states of the vortex system. The difference between these states is more quantitative than qualitative. Debye and Pearl shielding always occur in a system and do not replace one another during a transition to another state. The question is merely one of which phenomenon predominates. The arbitrariness of the KT transition in Pearl films is also related to this.

At intermediate defect concentrations, $n_{d1} < n_d < n_{d2}$, there is a range of temperatures within which the vortex system has two stable states [Fig. 1(b], of which the lower state corresponds to individual vortices and the upper, to collective vortices. Thus, a sufficiently high concentration of radiation defects in a superconductor will stabilize the collective state of the vortex system, since the number of trapped vortices in this state is determined primarily by the number of defects, as discussed above. Within this range of concentrations, hysteresis in the resistance that depends on the concentration of free vortices should be observed.

At very high defect concentrations $n_d > n_{d2}$, the left stability boundary of the metastable states shifts almost to $T = 0$. It can undergo a transition into the lower state only at very low temperatures. Figure 2 shows the temperature dependences of the concentrations of free and trapped vortices in this case. Almost all the defects in the film in the collective state have trapped a flux quantum apiece. Thus, the total concentration of vortices (free and trapped) is rather high, in order to ensure the collective behavior of the system, and depends weakly on the temperature. In this situation, as the temperature is lowered, the screening length δ begins to decrease, since it is proportional to $\sqrt{T}$. As a result, the creation of additional free vortices becomes energetically favorable at low temperatures, as Fig. 2 shows clearly. However, at sufficiently low temperatures, the collective state loses stability and the vortex system undergoes a transition to the individual state. The stability boundary of the collective state, which shows up here, is not an extension of the lower stability boundary into the region of medium defect concentrations discussed above. This can be understood by examining the phase diagram of the vortex system in the $n_d - T$ plane, which will be discussed in the next section.

All three scenarios have a common feature related to the presence of defects. For arbitrary defect concentrations, the free vortex concentration n, which is related to the resistive behavior of a superconductor, is higher than in a defect-free film and the jump in the resistivity is shifted toward lower temperatures.

This phenomenon can be explained as follows. When there are no radiation defects, the equilibrium concentration of free vortices develops because a dynamic equilibrium is established between the dissociation and recombination of vortex dipoles. Defects form an additional reservoir in which vortices can accumulate, since the state of a trapped vortex is more energetically favorable than that of a free vortex because of the core energy. This means that at low temperatures, when the system is in a state where collective effects are negligible, the concentration of trapped vortices is always higher than that of free vortices, while the concentration of free vortices is only slightly higher than in a defect-free film. As the temperature approaches T_{KT}, the concentrations of both free and trapped vortices increase, and this enhances the influence of collective effects on the state of the vortex system in both perfect and defective films. In the latter, however, the total vortex concentration is higher and collective effects make the jump in the vortex concentration occur at a lower temperature.

The numerical solutions presented here were obtained for a model film with the parameters $e_0 = 3$, $\Lambda/\xi = 10^2$ at $T = 0$, and $T_{c0}/T_{KT} = 1.2$. The critical values of the dimensionless defect concentration for these parameters, $n_{d1} \approx 0.03$ and $n_{d2} \approx 0.09$, are fairly high. The critical concentrations depend on these parameters and decrease slightly as the parameters are reduced.

4. PHASE DIAGRAM OF THE VORTEX SYSTEM

Knowledge of the $n_d - T$ phase diagram of a system of Pearl vortices in a superconducting film with radiation defects allows us to understand its behavior better. This diagram (Fig. 3) was constructed from a study of the minima in the free energy (3). It shows the phase transition curves separating the stability regions of the various states of the system.

The dashed curve *1–2* is the continuous ''phase transition'' curve. The changes in the concentrations of free and

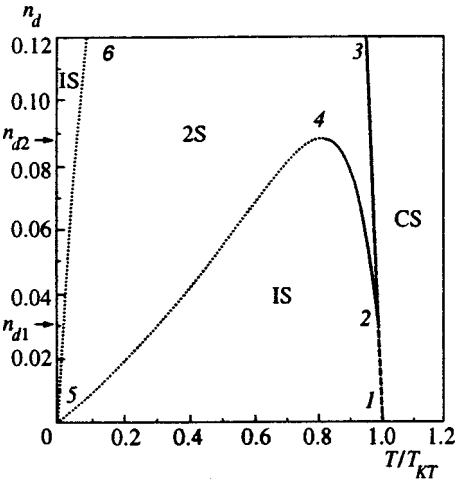

FIG. 3. $n_d - T$ phase diagram of a system of Pearl vortices in a thin superconducting film with radiation defects. The curves in this diagram separate the regions in which stable individual states (IS), stable collective states (CS), or two metastable states (2S) of the vortex system exist.

defect-trapped vortices on crossing this curve are shown in Fig. 1(a). To the left of it we have $\delta > \Lambda$ and the system is in the individual state (IS). To the right we have $\delta < \Lambda$ and the energy of the interaction between two vortices depends on the concentration of free vortices; this is the collective state of the system (CS).

Above the point *2* ($n_d > n_{d1}$), the phase transition curve splits into two. On the smooth curve *2–3*, the individual state of the system becomes stable, as well as the collective state. On the curve *2–4*, the collective state loses stability. For $n_d > n_{d2}$, the collective state remains stable at very low temperatures up to curve *5–6*, where the system undergoes a transition into the individual state. The temperature dependences of n and n_t within these ranges of the defect concentration n_d are plotted in Figs. 1(b) and 2.

Because of a weakening of the thermal fluctuations as the temperature is lowered, the collective state of the vortex system again becomes stable on the curve *4–5*, having lost stability on the curve *5–6*. A transition of the system into this state is improbable, since in the $n_d - T$ plane it can only move along a line parallel to the temperature axis. On any such line above n_{d1} the collective state is separated from the individual state by an energy barrier.

We conclude by examining the reliability of these results. The method used to obtain them is based on the London approximation for the vortex interaction energy and a gas approximation for calculating the partition function. Both approximations assume a low vortex concentration, i.e., $2n + 2n_t \ll 1$. This means that the method works poorly when the vortex concentration is such that the distance between them approaches the coherence length ξ. A situation of this sort arises in the region of the high temperature plateau in the $n(T)$ and $n_t(T)$ curves as $T > T_{\mathrm{KT}}$ and in the collective state of the system for $n_d > n_{d2}$. Here the dimensionless concentration of the free vortices approaches $n \sim 0.1$ and even exceeds it. Estimates show that the screening length for the vortex interaction in this region is comparable to and becomes shorter than the coherence length. Under these conditions, the magnetic interaction of the vortices ceases to predominate and it becomes necessary to take into account the interaction of their normal cores owing to the nonuniformity of the absolute value of the order parameter. In addition, a vortex system under these conditions should be regarded more as a liquid than as a gas. Thus, an accurate analysis of the system in this region will go beyond the framework of the London approach and employ a Ginzburg–Landau approximation, which is a much more complicated problem.

Based on the simple physical arguments given above, we may conclude that this approach yields a qualitative description of the behavior of a vortex system in these regions, as well. The existence of the curve *5–6* on which the collective state of the system loses stability is somewhat in doubt, since the reasons for this behavior are unclear. We have presented these results here, in order to illustrate all the answers that can be obtained using the model and techniques employed in this paper.

5. CONCLUSION

In this paper we have shown that radiation defects can have various effects on the properties of a thin superconducting film. On one hand, they cause pinning of magnetic vortices and a rise in the critical current for the transition of a superconductor into the resistive state. On the other hand, radiation defects form a reservoir in which vortices that are not bound in dipoles can accumulate and thereby expand the stability region for the collective state of the vortex system. In the collective state the system contains a significant number of free (not bound in dipoles and not trapped by defects) vortices, which are responsible for the resistive properties and reversible behavior of the superconductor.

The behavior of Pearl vortex systems described here allows us to reach several qualitative conclusions about the properties of superconducting thin films. The above discussion implies that, without an external magnetic field no significant increase in the critical current should be observed. In this case, the resistive behavior is associated, under the influence of the current with the motion of vortices that have entered from the edge of the sample. The current which detaches the vortices from the edge of the sample is of the same order of magnitude as the current that detaches vortices from defects. Thus, the behavior of defects should not radically change the situation. On the other hand, there is a drop in temperature at which the film begins to manifest resistive behavior owing to a transition into the collective state.

In an external magnetic field at low temperatures, an increase in the critical current can be observed owing to capture, by defects, of vortices which enter the sample under the influence of the field. After the onset temperature for avalanche growth in the vortex concentration is reached, the critical current should go to zero, since the number of free vortices increases rapidly. This is one of the possible mechanisms for depinning, which converts a superconductor into a resistive state and ensures reversible magnetization of the sample. In the case of a first-order transition, two irreversibility curves can be seen, depending on the direction of the change in the temperature during an experiment. A deeper

understanding of the physical processes leading to the appearance of an irreversibility curve will require an examination of the thermodynamics of a vortex system in a defective film placed in an external magnetic field and possibly a more accurate accounting for the effect of thermal fluctuations on the behavior of the system.

These comments apply equally to layered superconductors, since the mechanisms responsible for the resistive behavior of a sample and its magnetization are the same as for a Pearl film.

*)E-mail: artemov@kinetic.ac.donetsk.ua

[1] F. M. Sauerzopf, H. P. Wiesinger, H. W. Weber *et al.*, Physica C **162–164**, 751 (1989).
[2] R. B. Dover, E. M. Gyorgy, L. F. Schneemeyer *et al.*, Nature **342**, 55 (1989).
[3] L. Civale, A. M. Marwick, T. K. Worthington *et al.*, Phys. Rev. Lett. **67**, 648 (1991).
[4] Qiang Li, Y. Fukumoto, Y. Zhu *et al.*, Phys. Rev. B **54**, R788 (1996).
[5] C. J. van der Beek, M. Konczykowski, T. W. Li *et al.*, Phys. Rev. B **54**, R792 (1996).
[6] K. A. Muller, M. Takashige, and J. G. Bednorz, Phys. Rev. Lett. **58**, 1143 (1987).
[7] V. L. Berezinskiĭ, Zh. Éksp. Teor. Fiz. **61**, 1144 (1972) [Sov. Phys. JETP **34**, 610 (1972)].
[8] J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973).
[9] J. R. Clem, Phys. Rev. B **43**, 7837 (1991).
[10] K. H. Fisher, Physica C **178**, 161 (1991).
[11] J. Pearl, Appl. Phys. Lett. **5**, 65 (1964).
[12] M. R. Beasley, J. E. Mooij, and T. P. Orlando, Phys. Rev. Lett. **42**, 1165 (1979).
[13] R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics*, John Wiley and Sons (1975).
[14] V. N. Ryzhov and E. E. Tareyeva, Phys. Rev. B **49**, 6162 (1994).
[15] J. L. Chen and T. Y. Tseng, Phys. Rev. B **54**, 502 (1996)

Translated by D. H. McNeill

Soliton–magnon scattering in a two-dimensional isotropic magnetic material

B. A. Ivanov

Institute of Magnetism, National Academy of Sciences, 252142 Kiev, Ukraine; Taras Shevchenko Kiev University, 252127 Kiev, Ukraine

V. M. Murav'ev

Institute of Magnetism, National Academy of Sciences, 252142 Kiev, Ukraine

D. D. Sheka*)

Taras Shevchenko Kiev University, 252127 Kiev, Ukraine

(Submitted 12 April 1999)
Zh. Éksp. Teor. Fiz. **116**, 1091–1114 (September 1999)

We use the generalized σ-model to analytically study the solution of the problem of magnon scattering in two-dimensional isotropic ferromagnets and antiferromagnets in the presence of a Belavin–Polyakov soliton. We obtain the exact analytical solution to this problem for the partial mode with the azimuthal quantum number $m=1$. The scattering amplitude for other values of m (i.e., values not equal to unity) are studied analytically in the long- and short-wavelength approximations and also numerically for an arbitrary value of the wave number. We establish the general laws governing the soliton–magnon interaction. For a magnetic material of finite dimensions we calculate the frequencies of the magnon modes. We also use the data on local modes to derive the equations of motion of the soliton. Finally, we calculate the low-temperature (long-wavelength) asymptotic behavior of the magnon density of states due to the soliton–magnon interaction. © *1999 American Institute of Physics.* [S1063-7761(99)02309-4]

1. INTRODUCTION

It is now firmly established that solitons play an important role in low-dimensional magnetism, i.e., in one-dimensional (1D) and two-dimensional (2D) magnetic materials. Studies began with the simpler 1D case. Krumhansl and Schrieffer[1] found that solitons (kinks) must be considered on an equal basis with magnons as elementary excitations in the derivation of the thermodynamics of 1D magnetic materials. Currie *et al.*[2] construct a consistent phenomenological theory of solitons, in which a nontrivial fact was established, namely, that the kink–magnon interaction substantially alters the magnon density of states, which has an effect on the thermodynamic properties of the system. In particular, the temperature dependence of the soliton density is determined by the shift in the magnon phase in kink–magnon scattering and can vary substantially for magnetic materials with different kink–magnon interactions.[3,4]

A special role in soliton phenomenology is assigned to local magnon modes, which are spin waves localized at a magnetic soliton. For instance, the number of such modes determines the total variation of the magnon density of states and hence the temperature dependence of the kink density.[3] More than that, local modes are interesting objects by themselves, and their study is linked to direct experiments in exciting and detecting them, since by characterizing the intrinsic latent degrees of freedom of the soliton the local modes are the cause of soliton magnetic resonance at the characteristic frequencies of ''intrinsic'' motion.[5]

Important results in the soliton thermodynamics of 2D magnetic materials were obtained by Mertens *et al.*[6] and in the research that followed (see the review articles in Refs. 3 and 7). In research devoted to 2D solitons, the density of solitons (vortices) was taken as an external parameter of the theory. This approach was also used in analyzing the data of the experiments in which the contribution of localized 2D solitons to the relaxation of spin excitations was observed.[8–14] The main difficulty in analyzing 2D systems lies in the absence of exact analytical solutions for most models. Usually the solitons are treated numerically by diagonalizing with respect to small discrete systems.[15–18] In such finite geometry the soliton–magnon interaction manifests itself primarily in the existence of specific Goldstone local modes with anomalously low frequencies and in the excitation of magnon modes by soliton motion. Thanks to the reverse effect, it was possible to describe the dynamical parameters of a soliton by the data on local modes.[19]

In this connection, an important role is played by the analysis of such 2D models for which analytical results can be obtained and the general laws governing the soliton–magnon interaction can be established. Only one exact analytical solution of this type is known, the Belavin–Polyakov (BP) soliton, which describes a topological soliton in an isotropic 2D magnetic material.[20] The existence of local modes in such a system was predicted in Ref. 21 for an isotropic 2D ferromagnet and in Ref. 22 for an antiferromagnet. In particular, it was found that a BP soliton with a topological charge ν has $2|\nu|$ local modes of zero frequency (local zero-frequency modes).

In the present paper we construct a solution of the prob-

lem of scattering of magnons by a BP soliton in 2D magnetic materials. In Sec. 2 we examine the generalized σ-model, which can be used to describe ferromagnets and antiferromagnets, as well as ferrimagnets near the point of compensation of the sublattice spins. In Sec. 3 we formulate the scattering problem for this model and obtain its exact solution for the partial mode with azimuthal quantum number $m=1$. Sections 4 and 5 are devoted to calculations of the scattering amplitude for the other values of $m(n\neq 1)$ analytically in the long-wavelength approximation $kR\ll 1$ (Sec. 4) and numerically for arbitrary values of kR (Sec. 5), where k is the wave number and R is the radius of the soliton core. In the sections that follow we use the results to describe the various physical properties of solitons and local magnon modes. Section 6 deals with calculations of the frequency of the magnon modes for a magnetic material of finite dimensions. In the same section, using the data on local modes, we derive the equations of soliton motion. In Sec. 7 we calculate the magnon density of states for which the soliton–magnon interaction is responsible. In the Conclusion we discuss the different ways in which the theory could develop and the possible applications.

2. THE MODEL. ELEMENTARY EXCITATIONS

A broad class of classical isotropic Heisenberg 2D magnetic materials can be described dynamically in terms of the classical unit vector **n** of the order parameter, i.e., $n_z=\cos\theta$ and $n_x+in_y=\sin\theta\exp\{i\phi\}$. The dynamics of a classical ferromagnet is described by the Landau–Lifshitz equation for the normalized magnetization,[23] which acts as the dynamic variable **n**. In a classical antiferromagnet, the dynamic variable is the antiferromagnetism vector, which in the long-wavelength approximation can be assumed to be a unit vector. The dynamics of an antiferromagnet is described by the equations of the σ-model of the **n**-field.[24,25]

In the interests of generality we examine two types of magnetic materials within a unified approach, more precisely, on the basis of a generalized σ-model, whose Lagrangian in the 2D case can be written[26]

$$L=\frac{A}{2}\int d^2x\left\{\frac{1}{c^2}\left(\frac{\partial\theta}{\partial t}\right)^2-(\nabla\theta)^2+\sin^2\theta\left[\frac{1}{c^2}\left(\frac{\partial\phi}{\partial t}\right)^2-(\nabla\phi)^2\right]-\frac{2}{D}(1-\cos\theta)\frac{\partial\phi}{\partial t}\right\}, \tag{1}$$

where $A=JS^2$, where J is the exchange integral and S is the atomic spin. The specific type of magnetic material is determined by the relationship between the parameters c and D. To describe a ferromagnet we must drop the second time derivatives in the equations of motion, i.e., formally let c go to infinity. The dynamic term in the Lagrangian of the ferromagnet is of a purely gyroscopic nature, with the parameter D having the meaning of the spin stiffness of the ferromagnet. The dynamics of an isotopic σ-model describing an antiferromagnet has a Lorentz-invariant form with a characteristic speed parameter c. For an antiferromagnet there is no gyroscopic term (the coefficient D can be taken to infinity). Note that the generalized σ-model for finite D and c describes a ferrimagnet near the point at which the mechanical moments of the sublattices are balanced. For such a magnetic material the gyroscopic term has the same structure as in a ferromagnet but is proportional to the small parameter $(S_1-S_2)/(S_1+S_2)$, where S_1 and S_2 are the average mechanical moments of the sublattices.[27,28]

The simplest elementary linear excitations of a 2D isotropic magnetic material that arise against the background of the ground homogeneous state are the magnons belonging to the continuous spectrum. If we select the orientation of the order-parameter vector **n** along the polar axis, we get magnon solutions in the form of a circularly polarized wave $\theta=\text{const}\ll 1$, $\Phi=\mathbf{kr}-\omega(k)t$. The dispersion law for a ferromagnet is quadratic, $\omega_{FM}(k)=Dk^2$. For an antiferromagnet the dispersion law is linear, $|\omega_{AFM}(k)|=ck$, and there are two degenerate branches with opposite circular polarizations, $\omega=\pm ck$, which is equivalent to the possibility of linear polarization of magnons.

The simplest static nonlinear excitations in the 2D case are the BP solitons,[20]

$$\tan\frac{\theta_0}{2}=x^{-|\nu|},\quad \phi_0=\varphi_0+\nu\chi,\quad x=\frac{r}{R}, \tag{2}$$

which, naturally, has the same form for a ferromagnet, an antiferromagnet, and a ferrimagnet. Here r and χ are the polar coordinates in the plane of the magnetic material, the integer ν is the topological charge of the soliton, and R and φ_0 are arbitrary parameters.

The energy of such a soliton is given by the formula

$$E_0=4\pi A|\nu| \tag{3}$$

and is independent of R and φ_0. The ambiguity in the choice of φ_0 is a characteristic feature of many models and a consequence of the isotropy of the Heisenberg exchange. The existence of an arbitrary parameter R (the soliton radius) and the fact that the energy is independent of R are related to the scale invariance of the static two-dimensional σ-model.[23] Obviously, this symmetry is broken in dynamics, with the exception of the trivial case of a pure antiferromagnet and translational motion, when everything reduces to Lorentz transformations.

In analyzing the static solutions it is convenient to introduce the complex-valued order parameter $w=(n_x+in_y)/(1-n_z)$ and interpret it as a function of the complex variable $\zeta=re^{i\chi}$ describing the position of a point in the plane of the magnetic material. In terms of these variables, the static equations of the σ-model reduce to the self-duality equation[29] $\partial w/\partial\zeta=0$ or $\partial w/\partial\bar{\zeta}=0$. The BP soliton corresponds to the simplest solution of this equation of the form

$$w_0=A\zeta^{\nu}\ \text{ for }\ \nu>0,\quad w_0=A\bar{\zeta}^{-\nu}\ \text{ for }\ \nu<0. \tag{4}$$

There are also more general solutions to this equation of the form $=f(\zeta)$ or $w=f(\bar{\zeta})$, where f is any analytic function of the complex variable ζ. In particular, the static multisoliton solution with the topological charge ν depends on $2|\nu|$ parameters[23] and can be written

$$w=A\frac{\Pi_{k=1}^{\nu}(\zeta-a_k)}{\Pi_{k=1}^{\nu-1}(1-b_k\zeta)} \quad \text{for} \quad \nu>0 \tag{5}$$

(constructing the same general solution for $\nu<0$ is a trivial task). The energy associated with this solution is determined by (3) and is independent of the parameters A, a_k, and b_k. We associate the soliton center with the value $\theta=\pi$, this solution has (for different a_k and b_k) ν solitons with unit topological charges at the points $\zeta=a_k$. If all a_k coincide, then at $b_k=0$ the solution (5) coincides with (4) and describes one soliton with the topological charge ν at point $\zeta=a_k$. Thus, variations in the parameters a_k and b_k has a strong effect on the structure of the soliton but do not change its energy or topological charge. This means that a BP soliton has extremely high internal degeneracy, which reflects the property of conformal invariance of the static two-dimensional σ-model.[20,29] Hence a BP soliton consists of a set of local modes with a zero frequency. The explicit form of these zero-frequency modes can be obtained by varying (5) in the parameters a_k and b_k. In the limit $a_k, b_k \to 0$, the soliton can be represented by the expansion

$$\Omega\equiv\frac{w-w_0}{w_0}=\sum_{m=-\nu+1}^{\nu}\frac{A_m}{\zeta^m} \quad \text{for} \quad A_m\to 0, \tag{6a}$$

or, introducing the deviations θ and ϕ from the quantities θ_0 and ϕ_0 into the simplest equation (2), by the formula

$$\theta-\theta_0+i\sin\theta_0(\phi-\phi_0)=-\frac{\sin\theta_0 A_m}{(\bar{\zeta})^m}. \tag{6b}$$

This implies that there are $2|\nu|$ independent types of small perturbations that do not alter the soliton energy. Their form is determined by the function $\Omega\propto(\bar{\zeta})^{-m}\propto\exp\{im\chi\}$. This is equivalent to the statement that $2|\nu|$ local modes with a zero frequency are associated with a BP soliton (see below).

3. MAGNON MODES IN THE PRESENCE OF A SOLITON

To describe the magnon excitations that arise against the background of a BP soliton, it is convenient to introduce local coordinates $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ characterizing the distribution of the order parameter in a fixed soliton: $\mathbf{e}_3$ coincides with the order parameter $\mathbf{n}_0$ of the immobile soliton (2), $\mathbf{e}_1=\mathbf{e}_y\cos\phi_0-\mathbf{e}_x\sin\phi_0$, and $\mathbf{e}_2=\mathbf{e}_3\times\mathbf{e}_1$. Then the linear oscillations of the order parameter can be described in terms of the projections of $\mathbf{n}$ on the local axes $\mathbf{e}_1$ and $\mathbf{e}_2$: $\vartheta=\mathbf{n}\cdot\mathbf{e}_1$ and $\mu=\mathbf{n}\cdot\mathbf{e}_2$ (ϑ and $\mu/\sin\theta_0$ are the small deviations from θ_0 and ϕ_0, respectively).

The linearized equations for ϑ and μ can be represented in the form of the system of equations

$$\left[-\nabla_x^2+\frac{1}{x^2}\frac{\partial^2}{\partial\chi^2}+U_1(x)\right]\vartheta+\frac{2\nu}{x^2}\times\cos\theta_0\frac{\partial\mu}{\partial\chi}+\frac{R^2}{c^2}\frac{\partial^2\vartheta}{\partial t^2}+\frac{R^2}{D}\frac{\partial\mu}{\partial t}=0,$$

$$\left[-\nabla_x^2+\frac{1}{x^2}\frac{\partial^2}{\partial\chi^2}+U_2(x)\right]\mu-\frac{2\nu}{x^2}\times\cos\theta_0\frac{\partial\vartheta}{\partial\chi}+\frac{R^2}{c^2}\frac{\partial^2\mu}{\partial t^2}-\frac{R^2}{D}\frac{\partial\vartheta}{\partial t}=0, \tag{7}$$

where $\nabla_x^2\equiv(1/x)\partial/\partial x(x\partial/\partial x)$ is the radial part of the Laplace operator and $U_1(x)=(\nu/x)^2\cos 2\theta_0$ and $U_2(x)=\cot\theta_0\nabla_x^2\theta_0-(d\theta_0/dx)^2$ are the ''potentials.''[18,19] Using the explicit form (2) for the static solution, we can easily show that the ''potentials'' in both equations are the same. This fact is unique for the isotropic σ-model. For instance, for vortices in a magnetic material with easy-magnetization planes,[18,19] the potentials differ substantially. The very fact that the potentials are different not only complicates the analysis technically (in comparison to the ordinary Schrödinger equation) but also introduces serious problems. In particular, for systems of the form (7) with unequal potentials U_1 and U_2 many general assertions of the type of the oscillation theorem have yet to be formulated. In Ref. 18 it was shown that equations of this form may have truly localized states with an exponential decrease of the wave function and energies inside the continuous spectrum, which is forbidden for equations of the Schrödinger form.

In the degenerate case considered here the magnon modes can be described by a single complex-valued parameter $\Psi=\vartheta+i\mu$, which obeys the second equation

$$\left[-\nabla_x^2+\frac{1}{x^2}\frac{\partial^2}{\partial\chi^2}+\frac{\nu^2}{x^2}\cos 2\theta_0\right]\Psi-i\frac{2\nu}{x^2}\cos\theta_0\frac{\partial\Psi}{\partial\chi}+\frac{R^2}{c^2}\frac{\partial^2\Psi}{\partial t^2}-i\frac{R^2}{D}\frac{\partial\Psi}{\partial t}=0, \tag{8}$$

whose analysis is almost the same as that of the Schrödinger equation. It is convenient to seek the solution of Eq. (8) in the form of a partial-wave expansion:

$$\Psi=\sum_{m=-\infty}^{\infty}f_m\exp\{im\chi-i\omega t\}. \tag{9}$$

Here each partial wave f_n is an eigenfunction of the spectral problem

$$\hat{H}f_m=\varkappa^2 f_m, \quad \varkappa=kR, \tag{10}$$

for the 2D radial Schrödinger operator $\hat{H}=-\nabla_x^2+U_m(x)$ with the potential

$$U_m(x)=\frac{m^2+2m\nu\cos\theta_0+\nu^2\cos 2\theta_0}{x^2}.$$

The spectrum of the problem (10) is continuous and is described by functions of the form $f_m^{\varkappa}$, with $\varkappa\geqslant 0$. Clearly, the zero-frequency modes f_m^0 correspond to solutions[21]

$$f_m^{(0)}=x^{-m}\sin\theta_0. \tag{11}$$

These modes correspond to perturbations of the form (6), i.e., their presence is due to the conformal invariance of the problem. Here and below, for the sake of definiteness, we examine the case where $\nu>0$, and to analyze solitons with $\nu<0$ it is enough to replace m by $-m$. This solution be-

haves regularly as $r\to 0$ only for partial modes with $-\infty < m \leqslant \nu$. A simple analysis of Eq. (11) shows that for $-\nu+1 \leqslant m < \infty$ the function $f_m^{(0)}$, regular as $r\to 0$, also decreases far from the soliton. Hence for $-\nu+1 \leqslant m \leqslant \nu$ these functions are finite over the entire range of r. This corresponds to the earlier conclusion that a BP soliton with a topological charge ν has $2|\nu|$ local modes represented in (6). Note that the physical meaning of two of these modes is obvious: the translational mode $f_{m=1}^{(0)}$ describes the displacement of the soliton as a whole, and the rovibrational mode $f_{m=0}^{(0)}$ describes the rotation and change of the soliton radius (which corresponds to an ambiguity in the choice of the position of the soliton center and to arbitrary values of φ_0 and R, respectively). The established bound states (local modes) are limits for the magnon modes of the continuous spectrum as $\varkappa\to 0$, on contrast to the case of 1D magnetic materials (see the review in Ref. 3).

Using the standard method of varying the arbitrary constant, we can find the second linearly independent solution of Eq. (10) with $\varkappa=0$:

$$f_m^{(1)} = x^m\left(\frac{x^{2\nu}}{m+\nu}+\frac{2}{m}+\frac{x^{-2\nu}}{m-\nu}\right)\sin\theta_0, \tag{12}$$

which is regular at zero when $m>\nu$.

Thus, at $\omega=0$ one of the solutions, (11) or (12), for all m has no singularities at zero. We use this solution to analyze scattering at small (but finite) values of ω in the range of small r.

The exact solutions $f_m^{(0)}$ that have been found can be used to simplify the problem of the analysis of the continuous spectrum on the basis of the Darboux transformation.[30] The same approach has been used in Refs. 31 to study the 1D case. To explain the method, we introduce the Hermitian-conjugate lowering and raising operators

$$\hat{A} = -\frac{d}{dx}+\frac{f_m^{(0)\prime}}{f_m^{(0)}}, \quad \hat{A}^\dagger = \frac{d}{dx}+\frac{1}{x}+\frac{f_m^{(0)\prime}}{f_m^{(0)}}$$

such that $\hat{A}f_m^{(0)}=0$ (here and in what follows a prime stands for a derivative with respect to x). By introducing these operators we can represent the Schrödinger operator $\hat{H}$ in the factorized form $\hat{H}=\hat{A}^\dagger\hat{A}$. What is important is that this makes it possible to reformulate the initial problem (10) in terms of the eigenfunctions $g_m^\varkappa=\hat{A}f_m^\varkappa$ of the spectral problem of the form

$$\hat{\mathscr{H}}g_m=\varkappa^2 g_m, \quad \hat{\mathscr{H}}\equiv\hat{A}\hat{A}^\dagger=-\nabla_x^2+\mathscr{U}_m, \tag{13}$$

where the potential is

$$\mathscr{U}_m(x)=\frac{(m-1)^2+\nu^2+2\nu(m-1)\cos\theta_0}{x^2}.$$

Note that far from the soliton (as $\theta_0\to 0$) the potential $\mathscr{U}_m$ becomes the centrifugal potential of the form $(\nu+m-1)^2/r^2$, which depends explicitly on the azimuthal number labeled $m-1$, which explains the terminology used for the operators $\hat{A}$ and $\hat{A}^\dagger$.

The initial function f_m is restored by applying the raising operator:

$$f_m^\varkappa=\frac{1}{\varkappa^2}\hat{A}^\dagger g_m^\varkappa. \tag{14}$$

The transformation we have just carried out simplifies the problem for the translational mode ($m=1$) substantially. Indeed, in this case $\mathscr{U}_1(x)=\nu^2/x^2$ for all values of x, i.e., $g_1^\varkappa$ determines the free motion. In view of this, the regular solution has the form $g_1^\varkappa=-J_\nu(\varkappa x)$. Restoring the initial function by (14), we get

$$f_1^\varkappa = J_{\nu+1}(kr)-\frac{2\nu}{kr}\frac{J_\nu(kr)}{(r/R)^{2\nu}+1}. \tag{15}$$

The existence of this exact solution for all values of the wave vector $\mathbf{k}$ is a unique property of the model (1). For other values of m (as earlier, for the case of magnon scattering by magnetic vortices in ferromagnets[19] and antiferromagnets[18] with planes of easy magnetization), the problem can be solved only approximately or by numerical methods.

The solution (15) demonstrates an important feature of magnon modes, which is absent in the cases discussed in Refs. 18 and 19, where the exponential decrease of the deviation of magnetization from the easy-magnetization plane in a vortex far from the vortex center is a characteristic feature. Equation (15) shows that the deviation of $f_1^\varkappa$ from the asymptotic term $J_{\nu+1}(kr)$ is not localized in a region with a definite radius; instead it is characterized by a slower (power-law) decrease. More than that, for the most interesting cases of long-wavelength asymptotic behavior, for $k \ll 1/R$, the solution (15) over a broad range of values of r, or $R \ll r \ll 1/k$, has the same form as a combination of Bessel and Neumann functions, $J_{\nu+1}(z)\propto z^{\nu+1}$ and $N_{\nu+1}\propto z^{-(\nu+1)}$, i.e., the second term in (15) imitates the presence of the function N. (Below we will see that this property remains valid for all values of m.)

For magnetic vortices, the corresponding corrections are exponentially decreasing functions of the form $\exp\{-r/r_\nu\}$, where r_ν is the radius of the vortex core. In view of this, the scattering amplitude, i.e., the coefficient of the Neumann function, can be determined from the coefficient of $1/z^m$ in the region where $z\ll 1$. This is not true, however, in our case, with the result that the method developed in Ref. 19 for magnetic vortices and used to analyze the scattering matrix by analyzing the corrections to the zero-frequency modes in the region $r_\nu \ll r \ll 1/k$ needs to be thoroughly modified[1)] if we wish to use it in our problem. This modification is done in Sec. 4.

On the other hand, the terms with a power-law decrease of the form $1/r^a$, $a>1/2$, must be taken into account when we describe the properties of magnon modes in a magnetic material of finite size with a soliton. This is done in Sec. 6.

4. SCATTERING IN THE LONG-WAVELENGTH LIMIT

To describe the scattering of magnons by a BP soliton, we note that free magnon states can be found if we set $\nu=0$ in the ''potential'' $U_m(x)$. The resulting magnon modes $f_{m,\nu=0}^\varkappa = J_m(z)$, with $z=kr$, are the partial cylinder waves of a plane spin wave of the form

$$\exp\{i\mathbf{k}\cdot\mathbf{r}-i\omega t\}=\sum_{m=-\infty}^{\infty} i^m J_m(z)\exp\{im\chi-i\omega t\}. \quad (16)$$

In the presence of a soliton the behavior of the magnon solutions can be analyzed at large distances from the soliton ($r\gg R$). In view of the asymptotic behavior $U_m(x)\approx n^2/x^2$, in the leading approximation in $1/x$ we have the usual result[21]

$$g_m\propto G_m(z)\equiv J_{|n|}(z)+\sigma_m^{\nu}N_{|n|}(z), \quad n=\nu+m-1,$$

$$f_m\propto F_m(z)\equiv J_{|p|}(z)+\sigma_m^{\nu}N_{|p|}(z), \quad p=\nu+m \quad (17)$$

(below we also use the notation involving n and p, and $G_m(z)$ and $F_m(z)$ for the combination of cylinder functions of the specific form presented in (17) with allowance for σ). A comparison of the asymptotic behaviors of $G_m(z)$ and $F_m(z)$ with each other and with the solution (16) for free magnons suggests that $\sigma_m^{\nu}\equiv\sigma_m^{\nu}(\varkappa)$ determines the soliton–magnon scattering amplitude. Since the coefficients σ are the same for F_m and G_m, to calculate the scattering amplitude $\sigma_m(\varkappa)$ we can use the initial problem or the modified problem. In particular, there is no scattering for the translational mode. Unfortunately, there is no way in which we can find analytical solutions for the other modes, but the scattering can be analyzed fairly thoroughly in the limiting cases.

To analyze soliton–magnon scattering in the case of small k, we can use the fact that at $k=0$ we know the exact solutions f_m^0: (11) for $m\leqslant\nu$ and (12) for $m>\nu$. In this case, we can construct the solution for small but finite $k(k\ll 1/R)$ by using a perturbation-theory expansion in k^2. To this end we seek the solution in the form $f_m^{\varkappa}=f_m^0(1+\varkappa^2\alpha(x))$, where $\varkappa^2\alpha(x)\ll 1$. The function $\alpha(x)$ is determined by an inhomogeneous second-order linear equation, whose solution can be found by the method of variation of the arbitrary constant if the two linearly independent solutions of the homogeneous problem are known. For a magnetic material with easy-magnetization planes this can be done only for the translational mode.[19] In the case of an isotropic magnetic material the solutions can be found by this method for arbitrary values of m (see Refs. 32 and 22).

In deriving a specific solution it is convenient to employ the modified problem by using the first-order equation $\hat{A}^{\dagger}g_m^{\varkappa}=\varkappa^2 f_m^0$, where f_m^0 is the zeroth solution, bounded as $x\to 0$. When $m\leqslant\nu$, the function $f_m^{(0)}$ is such a solution, from which we easily find that

$$g_m^{\varkappa}=\frac{\varkappa^2}{xf_m^{(0)}}\Phi^{(0)}(x),$$

$$\Phi^{(0)}(x)=\int_0^x (f_m^{(0)}(\xi))^2\xi d\xi, \quad \text{for } m\leqslant\nu. \quad (18a)$$

The same formulas can be used to readily restore the explicit form of the solution $f_m^{\varkappa}$ of the initial problem:

$$f_m^{\varkappa}(x)=f_m^{(0)}(x)\left[1-\int_0^x\frac{g_m^{\varkappa}(\xi)}{f_m^{(0)}(\xi)}d\xi\right]. \quad (18b)$$

Analysis of this solution has shown that in a broad interval of r values, $0\leqslant r\ll R^s(1/k)^{1-s}$ (the values of parameter s are between 0 and 1 and depend on m), the addition to the zeroth solution $f_m^{(0)}$ is small and perturbation-theory techniques can be used.

The same laws stand for the magnon mode with $|m|=1$ scattered by a vortex in a magnetic material with easy-magnetization planes.[19] Since the deviations from the asymptotic solution were found to be exponentially small, both solutions are valid for $R\ll r\ll 1/k$, the asymptotic solution (18b) and of the form (17). This made it possible to find the coefficient of the Neumann function $N_{|n|}(kr)$ (with allowance for the fact that $N_{|n|}(kr)\propto(kr)^{-|n|}$ where $kr\ll 1$) and to write an analytical formula for $\sigma_{|m|=1}(k)$. In our case, however, the situation is more complicated. As noted in Sec. 3 in the discussion concerning the exact solution (15), the asymptotic solutions far from the soliton contains corrections that decrease by a power law. Although they do decrease faster than the asymptotic solution (18b), it is very important to account for them. In particular, they may have the same form as the Neumann function for $z\ll 1$.

Thus, to calculate the scattering amplitude we must compare the approximate solution (18b) not with the asymptotic form (17) but with the refined solution that allows for terms increasing by a power law far from the soliton. For $m\neq 1$ the corrections can be expressed by exact formulas, but they can easily be calculated in the long-wavelength approximation $k\ll 1/R$, where we can assume that $kR\ll z=kr\ll 1$.

To do this, we introduce the variable $z=kr=\varkappa x$ into Eq. (13). Then the combination $(R/r)^{\nu}$ in $\theta_0(r)$ becomes $\varkappa^{\nu}/z^{\nu}$ and vanishes for finite z as $\varkappa\to 0$. Hence in the limit $\varkappa=kR\to 0$ Eq. (13) simply becomes a Bessel equation with the solution (17), and the corrections can be found by a series expansion in powers of $(\varkappa/z)^{\nu}$. Keeping only the first nonvanishing approximation in $\varkappa^{\nu}$ and representing the asymptotic solution in the form $g_m^{\varkappa}(z)=G_m(z)+\mathcal{G}_{\nu,m}^{\varkappa}(z)$, we arrive at the inhomogeneous Bessel equation

$$\nabla_z^2\mathcal{G}+\left(1-\frac{n^2}{z^2}\right)\mathcal{G}=\frac{4\nu(1-m)}{z^2}\left(\frac{\varkappa}{2}\right)^{2\nu}G_m.$$

We see that to this accuracy the solution far from the soliton can be expressed in terms of the universal function $\mathcal{G}_{|n|\nu}(z)$,

$$g_m^{\varkappa}(z)=G_m(z)+4\nu(m-1)\varkappa^{2\nu}\mathcal{G}_{|n|,\nu}(z), \quad (19)$$

which can be found by solving an equation of the form

$$\nabla_z^2\mathcal{G}_{|n|,\nu}+\left(1-\frac{n^2}{z^2}\right)\mathcal{G}_{|n|,\nu}=-\frac{G_m}{z^{2(\nu+1)}}.$$

Using the standard method of variation of an arbitrary constant, we can write the solution of this equation in integral form:

$$\mathcal{G}_{|n|,\nu}(z)=\frac{\pi}{2}N_{|n|}\int_z^{\infty}\frac{G_m(z)J_{|n|}(z)}{z^{2(\nu+1)}}dz$$

$$-\frac{\pi}{2}J_{|n|}\int_z^{\infty}\frac{G_m(z)N_{|n|}(z)}{z^{2(\nu+1)}}dz. \quad (20)$$

Here integration can be carried out exactly, and at $|n|=\nu$ the answer is $\mathcal{G}_{\nu,\nu}=-g_{\nu}^{(0)}/8\nu^2z^{2\nu}$.

But if $|n|\neq\nu$, we have the recurrence relation

$$\mathscr{G}_{|n|,\nu}=\mathscr{G}_{\nu-1}A_{|n|,\nu}+B_{|n|,\nu},$$

$$A_{|n|,\nu}\equiv\frac{2\nu-1}{2\nu(n^2-\nu^2)},\quad B_{|n|,\nu}\equiv\frac{zg^{(0)}_{n-1}+(\nu-n)g^{(0)}_n}{4\nu(n^2-\nu^2)z^{2\nu}},$$

which yields

$$\mathscr{G}_{|n|,\nu}=\mathscr{G}_1\prod_{k=2}^{\nu}A_{|n|,k}+\sum_{k=2}^{\nu-1}B_{|n|,k}\prod_{i=k+1}^{\nu}A_{|n|,i}+B_{|n|,\nu}.$$

Limiting ourselves to corrections to the Bessel function, i.e., taking $G_m=J_{|n|}$ in (20), we arrive (after involved calculations that use the properties of cylinder functions) at an expression for $\mathscr{G}_1$:

$$\begin{aligned}\mathscr{G}_{|n|,1}(z)=&-\frac{J_{|n|}(z)}{4(n+1)}\left[\frac{1}{z^2}+\frac{\ln(z/2)-\psi(|n|+1)}{n(n+1)}\right]\\&+\frac{J_{|n|-1}(z)}{4z(n^2-1)}+\frac{1}{4|n|(n^2-1)}\\&\times\sum_{k=1}^{\infty}(-1)^k\frac{(|n|+2k)J_{|n|+2k}}{k(|n|+k)}\\&+\frac{\pi N_{|n|}(z)}{8|n|(n^2-1)},\end{aligned}\tag{21}$$

where $\psi(x)$ is the Euler psi function.

Thus, as with the exact solution (15), the asymptotic behavior of the solution for $r\gg R$ differs from that in magnetic materials with easy-magnetization planes discussed earlier. Even if $\sigma=0$, i.e., there is no scattering, in the region far from the solution but with r finite ($R\ll r\ll 1/k$), the solution contains a number of terms that formally diverge as $z\to 0$ ($kr\ll 1$). In this case, for $\sigma\neq 0$, in the region $R\ll r\ll 1/k$ of interest to us, the solution of the scattering problem can be written

$$g(z)\propto J_{|n|}(z)+4\nu(m-1)\varkappa^{2\nu}\mathscr{G}_{|n|,\nu}(z)+\sigma N_{|n|}(z).\tag{22}$$

Here we did not include the corresponding corrections to the Neumann function, since we can easily show that they contain higher orders of $\varkappa$ and are unimportant.

By comparing the approximate solution (18) valid for $0<x\ll 1/\varkappa$ with the solution (22) valid for $1\ll x\ll 1/\varkappa$ we can now find the scattering amplitude $\sigma_m(\varkappa)$. In analyzing this problem it is convenient to examine the different ranges of variation of the parameters separately.

1. The case $|n|<\nu$ incorporates both local modes with their numbers m taken from the interval $-\nu+1<m<1$ ($0<n<\nu$) and nonlocal modes for which $-2\nu+1<m<-\nu+1$ ($-\nu<n<0$). In finding the asymptotic solution (18a) of the equation we realize that far from the soliton the zero-frequency modes have the form

$$f^{(0)}_m\approx\frac{2}{x^{n+1}}\left(1-\frac{1}{x^{2\nu}}\right).\tag{23}$$

Hence we arrive at an approximation for $\Phi^{(0)}(x)$ in the important region $x\gg 1$:

$$\Phi^{(0)}(x)\approx\Phi_0-\frac{2}{n}x^{-2n}+\frac{4}{\nu+n}x^{-2n-2\nu},$$

$$\Phi_0=\frac{2\pi(\nu-n)}{\nu^2\sin(\pi n/\nu)},\tag{24}$$

where the contribution of the constant Φ_0 is crucial when $n>0$ but is a small correction when $n<0$.

Integration in (18a) with the use of (23) and (24) leads to an expression for $g^{\varkappa}_m$:

$$g^{\varkappa}_m(x)\propto\Phi_0x^n-\frac{2}{n}x^{-n}.\tag{25}$$

Here we have ignored terms of the form $x^{-2\nu}$ in comparison to $x^{-2|n|}$.

Let us compare the asymptotic solution we have just found with the solution (22) of the scattering problem. Using the expansion of cylinder functions for small values of z and comparing the resulting asymptotic expressions, we conclude that the term $x^{|n|}$ in (25) is related to the Bessel function $J_{|n|}$, while the term $x^{-|n|}$ is related to the Neumann function $N_{|n|}$ and determines the scattering amplitude σ. A simple comparison leads to an asymptotic expression for the scattering amplitude σ:

$$\sigma^{\nu}_m(\varkappa)=\frac{\pi(n\Phi_0/2)^{n/|n|}}{|n|!(|n|-1)!}\left(\frac{\varkappa}{2}\right)^{2|n|},\tag{26}$$

with $-2\nu+1<m<1$ and $m\neq-\nu+1$. Note that allowance for the corrections $\mathscr{G}_{|n|,\nu}$ leads to a contribution to σ of order $\varkappa^{2\nu}$, which is insignificant in the given range of parameters.

2. The case $|n|>\nu$ is realized for $m<-2\nu+1$ and $m>1$. Integration in (18a) with the use of the same approximations (23) and (24) leads to the asymptotic solution

$$g^{\varkappa}_m(x)\propto x^{|n|}\left(1+\frac{\nu+|n|}{\nu+n}x^{-2\nu}\right),\tag{27}$$

in which only the leading corrections in $1/x$ are retained.

The asymptotic expression (27) is valid for $m\leqslant\nu$, when the zeroth solutions f^0_m are described by the functions $f^{(0)}_m$. A similar calculation can be done for $m>\nu$, where for the zeroth functions we use $f^{(1)}_m$:

$$g^{\varkappa}_m=\frac{1}{xf^{(0)}_m}(1-\varkappa^2\Phi^{(1)}(x)),$$

$$\Phi^{(1)}(x)=\int_0^x f^{(0)}_m(\xi)f^{(1)}_m(\xi)\xi d\xi\quad\text{for}\quad m>\nu.\tag{28}$$

Calculating the integral in (28), we arrive at the asymptotic expression (27) for $\nu>1$. For $\nu=1$ the asymptotic solution for modes with $m>1$ is

$$g^{\varkappa}_m(x)\propto x^m\left(1+\frac{1}{x^2}-\frac{\varkappa^2}{m(m+1)}\ln x\right).\tag{29}$$

Thus, the asymptotic solutions (27) and (29) obtained for $|n|>\nu$ differ dramatically from the earlier solution (25): the solutions (27) and (29) do not contain terms of the form $1/x^{|n|}$ and hence cannot yield an asymptotic expression of the form $J_{|n|}+\sigma N_{|n|}$. This is possible only if in the solution (22) the correction $\mathscr{G}_{|n|,\nu}$ is balanced by the scattering term σN_n. Note that this is an extremely stringent condition: not only

must the terms $1/x^{|n|}$ be balanced but also all terms of the form $x^{2k}/x^{|n|}$, where $0 \leqslant k \leqslant |n|-1$. Allowing for the term in $\mathscr{G}_{|n|,\nu}$ related to N_n [see (21)], we get

$$4\nu(m-1)\varkappa^{2\nu}\frac{\pi N_n}{8n(n^2-1)}\prod_{k=2}^{\nu}A_k+\sigma N_n=0,$$

which yields a formula for the scattering amplitude:

$$\sigma_m^{\nu}(\varkappa)=\mathscr{A}_m^{\nu}\left(\frac{\varkappa}{2}\right)^{2\nu}, \quad -2\nu+1<m<1, \quad m\neq-\nu+1,$$

$$\mathscr{A}_m^{\nu}=-\frac{\pi 2^{\nu}(2\nu-1)!!}{(\nu-1)!|m|(m+1)\ldots(m+2\nu-1)}. \tag{30}$$

3. The special cases $|n|=\nu$ and $|n|=0$ where the solutions (30) and (26) become invalid include the translational mode ($m=1$), the local mode with $m=-\nu+1$, and the nonlocal mode with $m=-2\nu+1$. For the translational mode the exact solution (15) yields $\sigma=0$. A calculation done on the basis of (18a) and a comparison of the results with the solution (22) of the scattering problem lead in the other two cases to the following asymptotic expressions for σ:

$$\sigma_m^{\nu}(\varkappa)=\frac{\pi}{2\ln(1/\varkappa)}, \quad m=-\nu+1, \tag{31}$$

$$\sigma_m^{\nu}(\varkappa)=\frac{4\pi}{[(\nu-1)!]^2}\left(\frac{\varkappa}{2}\right)^{2\nu}\ln\frac{1}{\varkappa}, \quad m=-2\nu+1. \tag{32}$$

The above analysis of scattering in the long-wavelength limit makes it possible to calculate the scattering amplitude in the long-wavelength approximation, i.e., for $k \ll 1/R$. At this point in our discussion, several general remarks concerning the nature of soliton–magnon scattering are in order.

It was found that as $k \to 0$ the scattering amplitude $\sigma_m(k)$ tends to zero for all values of m and ν. In most cases the amplitude $\sigma_m(\varkappa)$ given by Eqs. (30) and (26) is a regular function of $\varkappa$. In contrast to (30) and (26), for parameter values specified by (31) and (32) there exists a derivative $d^p\sigma/d\varkappa^p$ that has a singularity. The order is $p=1$ for $m=-\nu+1$, with the scattering being at its maximum. Such nonanalytic behavior of $\sigma(k)$ was detected in the numerical analysis of scattering of magnons with $m=0$ by a vortex in an antiferromagnet with an easy-magnetization plane done in Ref. 18 (see also Ref. 32). The scattering intensity (in contrast to the case of magnetic vortices discussed in Refs. 18 and 19) is not at its maximum for partial waves with smallest values of m ($m=\pm 1,0$).

The very fact that for a partial wave with a given m the limit point $k=0$ serves as the local zero-frequency mode is not critical for the scattering intensity. In particular, the mode with $m=1$ (the well-known translational mode) does not undergo scattering.

We also note that for the case of scattering by a BP soliton there are no simple relationships that link the scattering intensities for $m=+|m|$ and $m=-|m|$. For scattering of magnons by a vortex in magnetic materials with easy-magnetization planes, such relationships were established by numerical analysis: for antiferromagnets $\sigma_m^{\nu}(k)=\sigma_{-m}^{\nu}(k)$ (Ref. 18), while for ferromagnets $\sigma_m^{\nu}(k)$ and $\sigma_{-m}^{\nu}(k)$ can be obtained from each other by changing the sign of the magnon frequency (Ref. 19)

In conclusion of this section, we give the general solution of the problem of the scattering of a plane spin wave by a BP soliton. It is convenient to formulate the solution in terms of the variable $\tilde{\Psi}=\Psi\exp\{i\nu\chi\}$, which becomes $(n_x+in_y)\exp\{-i\omega t\}$ as $r\to\infty$ and describes a spin wave propagating against the background of the homogeneous state with $\mathbf{n}\|\mathbf{e}_z$. The need to pass from Ψ to $\tilde{\Psi}$ can be explained by the fact that although far from the soliton the magnetization is homogeneous, $\mathbf{e}_3\|\mathbf{e}_z$, the unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ depend on χ. With allowance for (9) and (17), the asymptotic solution for $r \gg R$ can be written

$$\tilde{\Psi}=\sum_{m=-\infty}^{\infty}C_m(J_n(kr)+\sigma_m^{\nu}(\varkappa)N_n(kr))\exp\{in\chi-i\omega t\}, \tag{33}$$

where $n=\nu+m$, and the C_m are arbitrary constants. Using asymptotic expressions for the cylinder functions in the region $r \gg 1/k$ and selecting C_n on the basis of a comparison of (33) with the asymptotic expression (16) for free motion, we can write the general solution of the problem of scattering of a plane spin wave:

$$\tilde{\Psi}=\left[\exp\{i\mathbf{k}\cdot\mathbf{r}\}+\mathscr{F}(\chi)\frac{\exp\{ikr\}}{\sqrt{r}}\right]\exp\{-i\omega t\},$$

$$\mathscr{F}(\chi)=\frac{\exp\{-i\pi/4\}}{\sqrt{2\pi k}}\sum_{m=-\infty}^{\infty}(\exp\{2i\delta_m^{\nu}\}-1)\times\exp\{i(\nu+m)\chi\}. \tag{34}$$

In (34) we have introduced the scattering phase $\delta_m^{\nu}(\varkappa)$, which is related to the scattering amplitude by the simple formula $\sigma=-\tan\delta$.

The total scattering cross section is given by the formula

$$\varrho=\int_0^{2\pi}|\mathscr{F}|^2d\chi=\sum_{m=-\infty}^{\infty}\varrho_m,$$

where the $\varrho_m=(4/k)\sin^2\delta_m^{\nu}$ are the partial scattering cross sections.

As noted earlier, for small k, the maximum scattering is related to the local mode with $m=-\nu+1$, for which, according to (31), the scattering phase $\sigma=\pi/2\ln\varkappa$. Hence, in the leading approximation in k it is enough to limit oneself to the contribution of this mode, with the result that we arrive at an expression for the scattering function of the form

$$\mathscr{F}(\chi)\approx\sqrt{\frac{\pi}{2k}}\frac{\exp\{i(\chi+\pi/4)\}}{\ln kR}, \quad k \ll 1/R. \tag{35}$$

In this approximation the scattering is isotropic ($|\mathscr{F}(\chi)|$ is independent of χ). The corrections to this expression are of order $1/(kR)^{2\nu+1/2}$ and are important only for determining the anisotropy of $\mathscr{F}(\chi)$.

The total scattering cross section (which has an integrable singularity) in the limit $\varkappa\to 0$ is given by the formula

$$\varrho(\varkappa)\approx\frac{\pi^2}{k\ln^2 kR},\quad k\ll 1/R. \tag{36}$$

5. ANALYSIS OF SCATTERING DATA FOR MODERATE VALUES OF k

The scattering can also be treated analytically in the short-wavelength limit, $k\gg|m|/R$. It is natural to assume that in this case the problem can be analyzed in the quasiclassical approximation, which yields

$$g_m^{\varkappa}\propto\sqrt{\frac{p(x)}{x}}\cos\left[\text{const}+\int_{x_0}^{x}p(\xi)d\xi\right],$$

$$p^2(x)=\varkappa^2-\mathscr{U}_m(x)+\frac{1}{4x^2}. \tag{37}$$

Indeed, analysis shows that (37) is valid for all values of $z=kr$ larger than the coordinate of the turning point, $z_0=\varkappa x_0$, which corresponds to the condition $p(x_0)=0$. The value of x_0 is small, $x_0\sim|m|/\varkappa\ll 1$.

On the other hand, at small distances $r\ll R$ ($x\ll 1$) the ''potential'' $\mathscr{U}_m$ has the asymptotic form $\mathscr{U}_m\approx(\nu-m+1)^2/x^2$, i.e., it describes free magnons of the form (16) with a mixed index:

$$g_m\propto J_{|\nu-m+1|}(z),\quad f_m\propto J_{|\nu-m|}(z)\quad \text{for } r\ll R.$$

For $k\gg|m|/R$, there is a broad range of values of r, $|m|/k\ll r\ll R$, in which we can limit ourselves to the asymptotic expression for the Bessel function in the limit $z\gg 1$ and $z\gg|m|$:

$$g_m^{\varkappa}\propto J_{|\nu-m+1|}\approx\sqrt{\frac{2}{\pi z}}\cos\left(z-\frac{1}{2}|\nu-m+1|-\frac{\pi}{4}+\frac{4(\nu-m+1)^2-1}{8z}\right). \tag{38}$$

To within terms of order $1/z^2$, the solutions (37) and (38) coincide in the entire range of parameter overlap. Hence, doing the asymptotic expansion of (37) far from the soliton, we arrive at the short-wavelength asymptotic expression for the scattering amplitude:

$$\sigma_m^{\nu}(\varkappa)\approx\frac{\pi(m-1)}{\sin(\pi/2\nu)}\frac{1}{\varkappa},\quad \varkappa\gg|m|. \tag{39}$$

Most importantly, this formula reproduces a property of the exact solution (15) according to which $\sigma_m^{\nu}=0$ holds at $m=1$. More than that, the scattering amplitude asymptotically tends to zero as $1/\varkappa$ for all $m\neq 1$, with the σ being equal in absolute value but having opposite signs for magnon modes with $m=|m|$ and $m=-|m|+2$. Below we will see that this result plays an important role in the analysis of density of magnon states in a 2D magnetic material.

Now we can compare the scattering amplitudes in the long- and short-wavelength limits. Clearly, $\sigma\to 0$ in both cases, but the signs of $\sigma(\varkappa)$ for $\varkappa\to 0$ and $\varkappa\to\infty$ are opposite. This situation is characteristic of magnon scattering by a 1D soliton in the sine-Gordon and ϕ^4 models and of the Landau–Lifshitz equation (see the review article in Ref. 3). It can be assumed that for a certain finite $k=k_p$ the scattering amplitude has a pole. Naturally, there is no real divergence at this pole: the physically observed scattering phase δ_m^{ν} varies monotonically. The existence of a pole means that the total increment of the scattering phase, $\delta(\infty)-\delta(0)$, is finite. According to numerical calculations for a soliton with a topological charge $\nu=1$, this increment is equal to π (to within sign), i.e., each mode is associated with a single pole. Such a singularity manifests itself in the analysis of the number of magnon degrees of freedom (see Sec. 7).

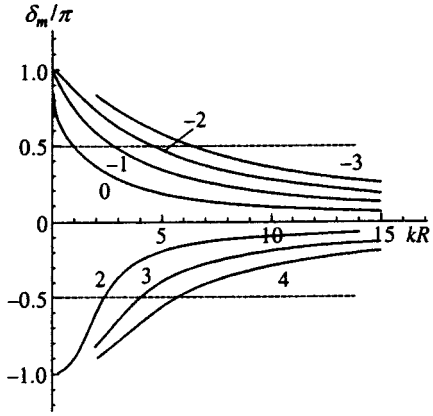

FIG. 1. Plots of δ_m vs. kR for $\nu=1$, labeled with the corresponding values of m. The dashed straight lines drawn through the values $|\delta_m|=\pi/2$ designate the positions of the poles of the scattering amplitude.

To analyze the intermediate values $kR\sim 1$, we solved the scattering problem numerically. The calculations were done by numerical integration of the spectral equations for the initial problem [Eq. (10)] and the modified problem [Eq. (13)] within a broad range of values of kR and m: $10^{-3}<kR<10^3$ and $-20\leqslant m\leqslant 20$ (the results of each calculation agree with what was said earlier). Basically we are interested in case with $\nu=1$, where the soliton energy is at its minimum. However, some data were obtained for $\nu=2$, 3, and 4, too.

Numerical calculations verified the long- and short-wavelength asymptotic expressions for the scattering amplitude given above. In the intermediate region of wave-vector values, $k\sim 1/R$, there are poles in the scattering amplitudes at $k=k+p$ for all the modes in question (Fig. 1 depicts the data for the modes with different values of m in the case of a soliton with $\nu=1$).

Let us discuss the problem of the position of the poles in the scattering amplitude in greater detail. According to the numerical data at $\nu=1$, for all $m\neq 1$ there is only one pole at $k=k_p$. Here k_p increases with $|m|$, and the functions $k_p=k_p(m)$ are different for $m=+|m|$ and for $m=-|m|$ (the reader will recall we are dealing with solitons with $|\nu|=1$). For very large values of $|m|$ the pole goes to infinity (Fig. 2). The situation becomes more complicated when $\nu>1$. More precisely, preliminary numerical data show that for a given m there can be several poles, with their number N_m not exceeding ν.

For comparative analysis of the scattering of modes with different values of m, we write explicitly the asymptotic expression for the scattering phase at $\nu=1$:

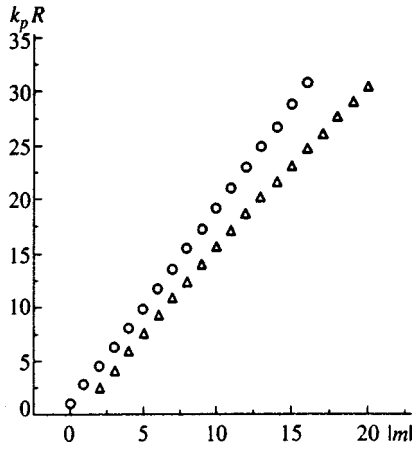

FIG. 2. The positions of the poles k_p as a function of the mode number m at $\nu=1$. The $\triangle$ correspond to $m>0$ and the $\bigcirc$ to $m\leqslant 0$.

$$\delta(\varkappa)\approx\begin{cases}\pi\,\mathrm{sgn}\,m\left(1-\dfrac{\varkappa^2}{2m(m+1)}\right), & \varkappa\ll 1, \quad m\neq -1,0,\\ \dfrac{\pi(1-m)}{\varkappa}, & \varkappa\gg|m|.\end{cases} \tag{40}$$

Assuming that these equations are valid at least qualitatively and setting $\varkappa\sim 1$, we can make a rough estimate of the position of the pole by equating the values of $\delta(\varkappa)$ for $\varkappa\ll 1$ and for $\varkappa\gg 1$. This yields $k_p\approx|m|/R$ for $|m|\gg 1$. Such an estimate reproduces fairly accurately the linear increase in k_p as a function of $|m|$ for large values of $|m|$ (see Fig. 2).

6. MAGNON MODES IN A MAGNETIC MATERIAL OF FINITE SIZE

The foregoing analysis of the scattering problem in the long-wavelength limit can be used to study the natural magnon modes in a magnetic material of a finite surface area containing a soliton. Such a problem plays an important role in many applications. Firstly, its solution can be used to describe analytically the data obtained through computer simulations of soliton motion, which are always done for systems of finite dimensions. In particular, in Refs. 33 and 19, this approach was used to describe the dynamics of a vortex in a ferromagnet with an easy-magnetization plane and to verify the non-Newtonian equations of motion containing third derivatives of the vortex coordinates with respect to time. Secondly, as noted earlier, this calculation can be used directly to describe the natural modes for the small particles of the magnetic material, which are in what is called the vortex state.[34]

We begin with the simplest case of the magnon modes in a circular system with a finite radius L and a soliton at the center. We discuss both the Dirichlet boundary conditions

$$\Psi(r,\chi)|_{r=L}=0, \tag{41}$$

which correspond to a fixed value of magnetization at the boundary, and the Neumann boundary conditions

$$\left.\frac{\partial\Psi(r,\chi)}{\partial r}\right|_{r=L}=0, \tag{42}$$

which model the case of free boundary conditions. There is no difficulty in extending these results to the case of general boundary conditions, but we do not do this here. The magnon spectrum in such a system is discrete. In the absence of solitons, the characteristic wave numbers $k_{m,i}$ are equal to $j_{m,i}/L$, where $j_{m,i}$ is the ith zero of the Bessel function J_m or the derivative of this function for the case (41) or (42), respectively.

In a magnetic material with a BP soliton, when k is large, we can ignore the local part of the function and write $J_n(kL)+\sigma(k)N_n(kL)=0$. It is natural, then, to expect the same behavior from k, i.e., $k=j/L$, where j lies between the values of the corresponding root of the Bessel or Neumann functions or the derivatives of these functions.

However, for $-\nu<m\leqslant\nu$, i.e., in the case of zero-frequency modes, the symmetry of the problem is high (scale invariance is restored). Hence we should expect the occurrence of Goldstone modes. In an unlimited (infinite) magnetic material, the frequencies of the Goldstone modes are zero, while in the presence of a boundary these modes manifest themselves as modes with very low frequencies, i.e., $kL\ll 1$. In particular, such modes arise for a vortex in a ferromagnet with an easy-magnetization plane in the case where $|m|=1$, which corresponds to translational motion of the vortex. For this mode, $k\sim r_v/L^2\ll 1/L$, where r_v is the radius of the vortex core. Since in this case the solution is approximated by (17) with an exponential accuracy even for $r>r_v$, the existence of Goldstone modes is determined solely by the scattering matrix.

When we are dealing with a BP soliton, in the analysis of Goldstone modes it is not enough to limit oneself to a solution in the form (17) corresponding to the scattering problem—one must also allow for the local part of the solution. The corresponding calculations are so tedious that in studying Goldstone modes it is more convenient to deal with the long-wavelength asymptotic expressions derived earlier; the expressions are valid for $r\ll 1/k$, i.e., for $kr<kL\ll 1$. It is this region that is so important in the analysis of such modes. Note that no Goldstone modes are present in the modified problem (the long-wavelength asymptotic expression (18a) has no small parameter, with the result that the boundary condition $g_m^{\varkappa}=0$ leads only to the solution $k\sim 1/L$).

In analyzing the Goldstone modes it is convenient to return to the initial problem for the function $f_m^{\varkappa}$. In this range of values of k, it is only natural to use the approximate expression (18b) for $f_m^{\varkappa}$. The analysis done using this expression shows that Goldstone modes occur only in the region where local modes exist. In the case of the Dirichlet boundary conditions, the spectrum of the Goldstone modes, which can be found from the condition $f_m(kL)=0$, has the form

$$kL=\begin{cases} 2\nu\sqrt{\dfrac{1+n}{\nu-n}\dfrac{\sin(\pi n/\nu)}{\pi}}\left(\dfrac{R}{L}\right)^n, & -\nu+1\leqslant m\leqslant\nu, \\ \sqrt{\dfrac{2}{\ln(L/R)}}, & m=-\nu+1. \end{cases} \tag{43}$$

The situation is somewhat more complicated for free boundary conditions. In particular, with Neumann boundary conditions, the solution (18b) does not allow for states with $kL\ll 1$. In this case, however, we can derive a solution by using a cylinder function of imaginary argument, which yields $\omega=Dk^2<0$ for the case of a ferromagnet or $\omega^2<0$ for an antiferromagnet. Below we discuss the physical meaning of negative values of ω and ω^2.

The following roots of the equation already agree with the condition $kL\sim 1$. They correspond to $k^2>0$ for all types of boundary conditions. Since for $k\sim 1/L$ and $R\ll L$ the ratio r/R is large at the boundary, the value of k_pL is close to the value of the corresponding zero of the Bessel function, $j_p\equiv j_{\nu+1,p}$, where $J_{\nu+1}(j_p)=0$, or to the value of the zero of the derivative, j_p', where $J_{\nu+1}(j_p')=0$ in the case of fixed and free boundary conditions, respectively:

$$k_pL=j_p+\frac{2\nu}{kL}\frac{J_\nu(j_p)}{J'_{\nu+1}(j_p)}\left(\frac{R}{L}\right)^{2\nu}, \tag{44a}$$

$$k_pL=J_p'+\frac{2\nu}{kL}\frac{(R/L)^{2\nu}}{J''_{\nu+1}(j_p')}\left\{\frac{2\nu+1}{kL}J_\nu(j_p')-J'_\nu(j_p')\right\}. \tag{44b}$$

Thus, the spectrum of the natural frequencies of a small particle of a magnetic material in an inhomogeneous state contains anomalously low frequencies, which manifest themselves in the magnetic resonance of samples containing such particles, say, ferroliquids and granular magnetic materials. Usov and Peschany[34] found that the magnetization distribution in a particle in the vortex state is well approximated by the BP soliton. Although our calculations can be applied only to particles shaped as a thin disk, it is easy to generalize them to the case of a cylinder.

Now we go back to the discussion of the meaning of the result $k^2<0$ for a Goldstone mode for free boundary conditions. We examine the most interesting case, $m=1$, corresponding to the translational motion of a BP soliton (below we will show that the parameters of a Goldstone mode can be directly related to the equations of motion of the soliton). The parameters of what is known as the translational Goldstone mode can be obtained directly from the exact solution (15). For $kR\ll 1$ the solution is

$$\Psi(x)\propto r^{\nu+1}\left(1-\frac{4\nu(\nu+1)}{(kr)^2}\frac{1}{(r/R)^{2\nu}+1}\right),$$

which implies that for fixed boundary conditions,

$$k^2=\frac{4\nu(\nu+1)}{L^2}\left(\frac{R}{L}\right)^{2\nu}. \tag{45}$$

For free boundary conditions the solution has the same form but k^2 is negative. Negative k^2 is not inconsistent with the presence of Bessel functions of imaginary argument (modified Bessel functions) in the solution, since we are studying this solution in the region $|k|\lesssim 1/L$, where the exponential increase of the function $I_n(z)\propto\exp\{z\}$ for $z\gg 1$ does not manifest itself.

For ferromagnets and antiferromagnets these results lead to significantly different physical pictures of soliton dynamics, which means that cases must be analyzed separately.

In the case of an antiferromagnet, there are two frequencies corresponding to the translational Goldstone mode:

$$\omega_0^2=\pm\frac{4\nu(\nu+1)c^2}{L^2}\left(\frac{R}{L}\right)^{2\nu}. \tag{46}$$

Clearly, this frequency has meaning only for fixed boundary conditions, and negative ω^2 mean that the system is unstable. At the same time, for a ferromagnet the value $\omega=Dk^2<0$ does not contradict the condition for stability. These results can easily be explained on the basis of a simple physical picture of soliton motion.

Obviously, for an antiferromagnet, which is described by Lorentz-invariant equations, the dynamics of all excitations must also be Lorentz-invariant. When the soliton is slow, $v\ll c$, this means that in the leading approximation the soliton coordinate $\mathbf{X}$ (for $\mathbf{X}$ the origin is at the center of the system) in the case of an antiferromagnet satisfies an equation of the Newtonian type:

$$M\frac{\partial^2\mathbf{X}}{\partial t^2}=\mathbf{F}_e, \tag{47a}$$

where $\mathbf{F}_e$ is the external force acting on the soliton, and $M=E_0/c^2$ is the effective soliton mass, with E_0 the soliton energy [see Eq. (3)]. Assuming that when the deviation of the soliton from the equilibrium position at the center of the system is small we can write

$$\mathbf{F}_e=\frac{\alpha\mathbf{X}}{L^p}, \tag{47b}$$

let us compare the value of the frequency obtained by (46) with the value of $\omega^2=-\alpha/ML^p$. We find that $p=2(\nu+1)$, and $\alpha=\mp 16\pi\nu^2(\nu+1)AR^{2\nu}$ for the Dirichlet and Neumann boundary conditions, respectively. This corresponds to the simple picture according to which $\mathbf{F}_e$ is the force of the image acting on the soliton because of the presence of a boundary. Since magnetic vortices interact as 2D charges and a BP soliton with $\nu=1$ is a vortex dipole, solitons with given $\nu>1$ can be interpreted as 2ν-multipoles, which explains the presence of p in (47b) and the sign of α.

Thus, the properties of the translational Goldstone mode in an antiferromagnet can easily be understood from the following reasoning. When a soliton is deflected from its equilibrium position $\mathbf{X}=0$, it is driven by the force of the image. For the Dirichlet boundary conditions the force is a restoring one (repulsion from the boundary) and the motion is stable. If the soliton is attracted to the boundary (the Neumann boundary conditions), Eq. (47) describes the departure of the soliton from the unstable position of equilibrium at $\mathbf{X}=0$.

Allowance for the next values of $k_{n,i}$ for $i>0$ can also be explained on the basis of effective equations for $\mathbf{X}$. Here the hierarchy of the effective equations of motion containing only even-order time derivatives manifests itself. The coef-

ficients of the higher-order derivatives diverge as $L \to \infty$. Mertens *et al.*[33] proposed equations of this type for describing the behavior of interplanar vortices in a ferromagnet.

The situation is quite different for a ferromagnet. The equation that is commonly used to describe the soliton dynamics is

$$M \frac{\partial^2 \mathbf{X}}{\partial t^2} + G\left(\mathbf{e}_z \times \frac{\partial \mathbf{X}}{\partial t} \right) = \mathbf{F}_e . \tag{48}$$

Here $\mathbf{F}_e$ is the external force, which, obviously, is the same as in the case of an antiferromagnet [see Eq. (47b)], and G is the gyroscopic term, whose value is determined only by the topology and has been reliably established,[3,24,35,36] $G = 4\pi\nu A/D$. The data on the effective mass of 2D solitons and vortices are contradictory: in Ref. 37 it is stated that in a ferromagnet with an easy-magnetization plane the value of M is finite but diverges as the anisotropy constant K tends to zero, $M \propto 1/K$. In Ref. 38 the result for a vertex is $M \propto 1/L$, in Ref. 33 the mass M is proportional to $\ln L$, and in Ref. 19 M is finite, but only if the term $G_3(\mathbf{e}_z \times \partial^3 \mathbf{X}/\partial t^3)$ is present in the effective equations of the form (48). In Refs. 36 and 39, the dynamics of a BP soliton is described on the basis of the Hamiltonian formalism with noncanonical Poisson brackets, and the relationship between momentum and velocity and the values of the mass are not discussed.

According to (45), the frequency of the translational Goldstone mode for a ferromagnet has the form

$$\omega_0 = \pm \frac{4\nu(\nu+1)D}{L^2} \left(\frac{R}{L} \right)^{2\nu}, \tag{49}$$

with the ''plus'' and ''minus'' corresponding to the Dirichlet and Neumann boundary conditions, respectively. In the present case there is no instability, since Eq. (48) with $M = 0$ (i.e., only the gyroscopic term is taken into account) describes small oscillations of the soliton in the case of attraction to the boundary and in the case of repulsion from the boundary. Allowance for the next translation mode, whose frequency is determined by the formula

$$\omega_1 = D(j/L)^2 \quad \text{or} \quad \omega_1 = -D(j'/L)^2 \tag{50}$$

in the case of the Dirichlet or Neumann boundary conditions, respectively, makes it possible to draw a conclusion about the inertial terms in the equation of motion.

Assuming that $\omega_0 \ll \omega_1$, these roots can easily be compared with the two frequencies that arise in the solution of Eq. (48). Indeed, in this case we have $\omega_0 \approx -\alpha/GL^p$, which yields exactly the first value of the frequency of the translational Goldstone mode. For the second value we get $\omega_1 \approx -G/M$. This value can be compared to (50) if we put

$$M = -\frac{4\pi\nu A}{D^2} \left(\frac{L}{j} \right)^2 \quad \text{or} \quad M = \frac{4\pi\nu A}{D^2} \left(\frac{L}{j'} \right)^2, \tag{51}$$

respectively, for fixed or free boundary conditions. Thus, as for a vortex, the dynamics with the frequency ω_1 is determined by the entire region to which the magnetic material is confined. Just as the coefficient G_3 in the third-order equations for vortices in a ferromagnet is nonlocal, so is the coefficient M: it depends on the boundary conditions and diverges as $L \to \infty$. The divergence of M is probably a general property of 2D magnetic materials with a gapless dispersion law.

We also note that the finite value of the soliton mass $M \propto 1/K$, where K is the anisotropy constant, obtained in Ref. 37 for a magnetic material with an easy-magnetization, axis does not contradict the above dependence $M \propto L^2$ for an isotropic ferromagnet. Indeed, in a magnetic material with an easy-magnetization axis, the gap in the magnon spectrum is finite and a characteristic linear scale $\Delta_0 = \sqrt{A/K}$ appears, from which we can obtain the same result as in Ref. 37, $M \propto \Delta_0^2 \propto 1/K$, if L is replaced by Δ_0 in (51).

7. DENSITY OF MAGNON STATES OF A 2D ISOTROPIC MAGNETIC MATERIAL IN THE PRESENCE OF A SOLITON

A 2D magnetic material can be described thermodynamically with allowance for soliton excitations via a generalization of soliton phenomenology developed by Krumhansl and Schrieffer[1] and Currie *et al.*[2] for 1D systems to the two-dimensional case. According to their approach, at low temperatures the state of a 1D magnetic material can be described in terms of almost free excitations, magnons and kinks. The main effect of their interactions manifests itself in the form of an asymptotic shift of the phase of a magnon scattered by a kink. This causes the total number of magnon states from the continuous spectrum to change (in comparison to the case of a magnetic material without a soliton) by $\Delta N = \int_{-k_0}^{k_0} \rho(k)\,dk$, where $\rho(k) = (1/2\pi)\,d\delta(k)/dk$ is the density of states. This quantity is a negative integer, i.e., the number of magnon states in the presence of a soliton decreases by ΔN, which is obvious, since a fraction of the magnon states are now described as the collective modes of the kink dynamics. The variation of the density of magnon states due to the addition of a kink to the system causes a change in the thermodynamic characteristics of the magnon gas, in particular, the free energy of the magnons. In the phenomenological approach, this change in the free energy of magnons is interpreted as a change in the kink energy due to kink–magnon interaction.

Let us use all these ideas in the 2D case. Clearly, in a 2D magnetic material the total number of states is proportional to $L_x L_y$. A free magnon corresponds to the expansion (16) in the cylinder harmonics $J_m(kr)\exp\{im\chi\}$ in which the angular variable has already been quantized, so that only the radial part $J(kr)$ needs to be quantized. In a circular geometry with radius L, the simplest quantization condition (41) has the form $J_m(kL) = 0$, from which it follows that $k_n L = j_{m,n}$. In the region of interest to us, $n \gg 1$, the zeros of the Bessel functions, $j_{m,n}$, are approximately equal to πn. From this fact we can formally determine the admissible values of the wave number by the same expression as in the 1D case. However, one must bear in mind that such an approximation for $j_{m,n}$ is valid only when m is not very large. For modes with $|m| \gg 1$ the first zero $j_{m,1} \approx |m|$. Hence in a system whose size L is finite there is a restriction on the admissible numbers of the modes, namely, $|m| \leqslant L$. Allowing for this

fact, we arrive at a rule for summing over the magnon states for a 2D magnetic material without a soliton:

$$\sum_{k,m} = \frac{L}{\pi}\int_0^{k_0} dk \sum_{m=-kL}^{kL}.$$

Naturally, for the total number of magnon states we arrive at the usual formula $N_{2D}=L^2k_0^2/\pi$.

Allowance for the soliton–magnon interaction leads to a shift in the magnon phase and changes, just at its does in 1D systems, the expression for the density of states (in our case, partial states for magnons with a given m) $\rho_m(k) = (1/\pi)d\delta_m(k)/dk$. The total density of magnon states is found by summing over m:

$$\mathscr{R}(k) = \sum_{m=-kL}^{kL} \rho_m(k) = \frac{1}{\pi}\sum_{m=-kL}^{kL} \frac{d\delta_m(k)}{dk}. \tag{52}$$

Note that the density of states $\mathscr{R}(k)$ in the long-wavelength region has an (integrable) divergence caused by the mode with $m=0$, for which, according to (31), $\rho_0(k) \approx (2k)^{-1}\ln^{-2}(kR)$ diverges in the limit $kR\to 0$ [cf. (36)]. It is also obvious that at low temperatures, $T \ll T_*$, where $T_* = \hbar D/R^2$ for ferromagnets and $T_* = \hbar c/R$ for antiferromagnets, it is enough to limit oneself to the long-wavelength approximation. In particular, in the adopted approximation, the density of energy states can be written

$$g(E) \propto \frac{1}{ER\ln^2(E/T_*)}. \tag{53}$$

In principle, the density of states for an arbitrary k can be calculated numerically. Here the presence of a pole in the amplitude of scattering of magnons with a given m means that the total phase changes by $+\pi$ or by $-\pi$ as k changes from zero to infinity, with the modes with $m>1$ and $m<1$ providing contributions to $\mathscr{R}(k)$ that are opposite in sign. Thus, for values of k that are not small the total number of magnon states does not decrease (as it does in the 1D case); rather, the magnon modes are redistributed among the states with different values of m. In general the signs in the series (52) are found to alternate. In thermodynamic calculations the temperature acts as a sort of regularizing factor in this summation process. The main contribution of the various modes, in particular, the change of the number of partial states by one unit, manifests itself in the order in which the poles k_p appear in the scattering amplitude as k increases. Since k_p increases with m (see Sec. 5), the contributions of the modes with an ever increasing m manifest themselves successively as the temperature rises.

8. CONCLUSION

Thus, we have constructed the soliton–magnon scattering matrix for the simplest but physically interesting 2D model of an isotropic magnetic material. The analysis has been carried out both for the Landau–Lifshitz equation, used to describe ferromagnets, and for the Lorentz-invariant σ-model, used in field theory and to describe antiferromagnets. We are the first to obtain an exact solution of the scattering problem for the partial mode with the azimuthal quantum number $m=1$. Note that such solutions are not known for all one-dimensional problems.

What is important is that the possibility of such an investigation is not related to exact integrability of the problem. Indeed, the model of an isotropic magnetic material is exactly integrable in the static case, $\mathbf{n}=\mathbf{n}(x,y)$, but nothing is known of its integrability in the case $\mathbf{n}=\mathbf{n}(x,y,t)$.

We have calculated the scattering amplitude for $m\neq 1$ (analytically in the long-wavelength approximation $kR\ll 1$ and for large values of kR and also numerically for arbitrary values of kR). We have found that the partial scattering amplitudes have poles (the scattering phases pass through $\pi/2$) at certain values $k=k_p$, with k_p increasing with m approximately by a linear law. This is enough to calculate the magnon density of states in the presence of a soliton.

We have used our results to describe various physical properties of solitons and local magnon modes. In particular, we have calculated the frequencies of the magnon modes for a magnetic material of finite dimensions. What we have found is that in the small particles of ferromagnets containing a soliton (particles in what is known as the vortex state, whose properties are being widely discussed at present) natural modes arise with anomalously low frequencies. The data on the frequencies of the local modes have been used to derive the equations of motion of a soliton in a ferromagnet. We have calculated the magnon density of states in the presence of a soliton, which makes it possible to construct a soliton phenomenology for 2D magnetic materials that allows for the soliton–magnon interaction.

There are other possible applications of our results worth noting. In some of the papers (see, e.g., the review article in Ref. 7) devoted to the study of ordered 1D media including magnetic materials, several nonequilibrium characteristics of a soliton gas, primarily, the coefficients of diffusion and viscosity, were investigated. The theories developed by the researchers were based on using the exact wave functions of magnons against the background of a soliton. The asymptotic expressions for the wave function for small k derived in the present paper have made it possible to study the irreversible process for the 2D gas of elementary excitations, including solitons and magnons, in isotropic magnetic materials at low temperatures.

The results concerning the σ-model can easily be extended to the Euclidean case and can be used to describe the quantum properties of spin chains with antiferromagnetic interaction. The properties of such systems are determined by the instantons of the Euclidean version of the nonlocal σ-model. Also widely discussed are instantons with a structure of the BP soliton (see Ref. 40) and what is known as merons, which have a half-integer topological charge (see Ref. 41). To calculate the pre-exponential factors in the corresponding transition amplitudes (the fluctuation determinant), we must know the complete set of eigenstates against the instanton background. Most important are zero-frequency modes (for more details see Ref. 42). Hence our results, especially concerning the nontrivial local zero-frequency modes, may prove to be important in developing the instan-

ton approach in the quantum theory of 1D magnetic materials.

The authors are grateful to V.G. Bar'yakhtar, G. Holzwarth, F.G. Mertens, D.I. Sheka, and G.M. Wysin for discussing the results. The work was partially sponsored by the Ukrainian Foundation for Basic Research (Project 2.4/27). The research of one of the authors (D.D.Sh.) was sponsored by the International Sciences and Education Program (Grant No. YSU082065).

*)E-mail: sheka@rpd.univ.kiev.ua

1)If this fact is not taken into account, the amplitude for scattering of the translational mode by a BP soliton turns out to be finite,[32,22] whereas according to (15) it must be zero.

[1] J. A. Krumhansl and J. R. Schrieffer, Phys. Rev. B **11**, 3535 (1975).
[2] J. R. Currie, J. A. Krumhansl, A. R. Bishop, and S. E. Trullinger, Phys. Rev. B **22**, 477 (1980).
[3] B. A. Ivanov and A. K. Kolezhuk, Fiz. Nizk. Temp. **21**, 355 (1995) [Low Temp. Phys. **21**, 275 (1995)].
[4] M. M. Bogdan and A. S. Kovalev, Z. Phys. B: Condens. Matter **71**, 341 (1988).
[5] J.-P. Boucher, G. Rius, and Y. Henry, Europhys. Lett. **4**, 1073 (1987).
[6] F. G. Mertens, A. R. Bishop, G. M. Wysin, and C. Kawabata, Phys. Rev. B **39**, 591 (1989).
[7] V. G. Bar'yakhtar and B. A. Ivanov, in *Soviet Scientific Reviews, Section A*, I. M. Khalatnikov (Ed.), Vol. 16, No. 3 (1993), p. 1.
[8] F. Waldner, J. Magn. Magn. Mater. **31–34**, 1203 (1983).
[9] F. Waldner, J. Magn. Magn. Mater. **54–57**, 873 (1986).
[10] F. Waldner, J. Magn. Magn. Mater. **104–107**, 793 (1992).
[11] C. E. Zaspel, Phys. Rev. Lett. **48**, 926 (1993).
[12] C. E. Zaspel, T. E. Grigereit, and J. E. Drumheller, Phys. Rev. Lett. **74**, 4539 (1995).
[13] C. E. Zaspel and J. E. Drumheller, Int. J. Mod. Phys. **10**, 3649 (1996).
[14] L. Subbaraman, C. E. Zaspel, and J. E. Drumheller, Phys. Rev. Lett. **80**, 2201 (1998).
[15] G. M. Wysin, Phys. Rev. B **49**, 8780 (1994).
[16] G. M. Wysin and A. R. Völker, Phys. Rev. B **54**, 12921 (1996).
[17] G. M. Wysin, Phys. Rev. B **54**, 15156 (1996).
[18] B. A. Ivanov, A. K. Kolezhuk, and G. M. Wysin, Phys. Rev. Lett. **76**, 511 (1996).
[19] B. A. Ivanov, J. J. Schnitzer, F. G. Mertens, and G. M. Wysin, Phys. Rev. B **58**, 8464 (1998).
[20] A. A. Belavin and A. M. Polyakov, JETP Lett. **49**, 245 (1975).
[21] B. A. Ivanov, JETP Lett. **61**, 917 (1995).
[22] B. A. Ivanov, V. M. Murav'ev, and D. D. Sheka, Ukr. Fiz. Zh. **44**, 500 (1999).
[23] A. M. Kosevich, B. A. Ivanov, and A. C. Kovalev, *Nonlinear Magnetization Waves. Dynamic and Topological Solitons* [in Russian], Naukova Dumka, Kiev (1983).
[24] I. V. Bar'yakhtar and B. A. Ivanov, Fiz. Nizk. Temp. **5**, 759 (1979) [Sov. J. Low Temp. Phys. **5**, 361 (1979)].
[25] A. F. Andreev and V. I. Marchenko, Usp. Fiz. Nauk **130**, 39 (1980) [Sov. Phys. Usp. **23**, 21 (1980)].
[26] B. A. Ivanov and D. D. Sheka, *Chaos, Solitons, & Fractals* (Special Issue ''Solitons in Science and Engineering: Theory and Applications''), Vol. 5, (1995), p. 2605.
[27] B. A. Ivanov and A. L. Sukstanskiĭ, Zh. Éksp. Teor. Fiz. **84**, 370 (1983) [Sov. Phys. JETP **57**, 214 (1983)].
[28] B. A. Ivanov and A. L. Sukstanskiĭ, Solid State Commun. **50**, 523 (1984).
[29] A. M. Perelomov, Usp. Fiz. Nauk **134**, 577 (1981) [Sov. Phys. Usp. **24**, 645 (1981)].
[30] V. Matveev and M. Salle, *Darboux Transformations and Solitons*, Springer-Verlag, New York (1991).
[31] G. E. Khodenkov, Fiz. Met. Metalloved. **61**, 850 (1986); P. Sodano, M. El-Batanony, and C. R. Willis, Phys. Rev. B **34**, 4936 (1981); B. A. Ivanov and H. Benner, J. Phys.: Condens. Matter **11**, 513 (1999).
[32] B. A. Ivanov and V. M. Murav'ev, Fiz. Nizk. Temp. **24**, 672 (1998).
[33] F. G. Mertens, H. J. Schnitzer, and A. T. Bishop, Phys. Rev. B **56**, 2510 (1997).
[34] N. A. Usov and S. E. Peschany, J. Magn. Magn. Mater. **118**, L290 (1993).
[35] A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials*, Applied Solid State Science Series, Supplement I, Academic Press, New York (1979).
[36] S. Komineas and N. Papanicolaou, Physica D **99**, 81 (1991).
[37] B. A. Ivanov and V. A. Stephanovich, Phys. Lett. A **141**, 89 (1989).
[38] F. G. Martens *et al.*, in *Nonlinear Coherent Structures in Physics and Biology*, F. G. Mertens and K. H. Spatschek (Eds.), Plenum Press, New York (1994).
[39] N. Papanicolaou and T. N. Tomaras, Nucl. Phys. **360**, 425 (1991).
[40] E. Fradkin, in *Frontiers in Physics*, Vol. 82, Addison-Wesley, Reading, MA (1991).
[41] Ian Affleck, J. Phys.: Condens. Matter **1**, 3047 (1989).
[42] A. I. Vaĭnshteĭn, V. I. Zakharov, V. A. Novikov, and M. A. Shifman, Usp. Fiz. Nauk **136**, 553 (1982) [Sov. Phys. Usp. **25**, 195 (1982)].

Translated by Eugene Yankovsky