Local spectroscopy of carrier reflection at the interface between a normal metal and the Peierls conductor $K_{0.3}MoO_3$

A. A. Sinchenko

Moscow State Engineering-Physics Institute, 115409 Moscow, Russia

Yu. I. Latyshev, S. G. Zybtsev, and I. G. Gorlova*)

Institute of Radio Technology and Electronics, Russian Academy of Sciences, 103907 Moscow, Russia
(Submitted 30 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1830–1842 (May 1998)

The nature of carrier reflection from a normal metal–Peierls conductor interface is clarified by studying the characteristics of point contacts formed with an intermediate metallic layer sputtered onto the (010) face of $K_{0.3}MoO_3$ crystals (injection along the chains) or the $(\bar{2}01)$ face (injection perpendicular to the chains). In the Peierls state, for bias voltages eV smaller than the Peierls gap Δ_P, an excess differential resistance is observed with a local minimum at $V=0$. The magnitude of the excess resistance is proportional to a^2/d^2, where a is the contact diameter and d is the thickness of the metal film. The excess resistance is much higher for injection along the chains than for injection perpendicular to them. A comparative analysis of the data for different injection directions indicates that the dominant contribution to the excess resistance for injection along the chains is from normal reflection of carriers without changes in the sign of their charge and with a momentum transfer $2p_F$ to the condensate of electron-hole pairs carried away from the interface. © *1998 American Institute of Physics.* [S1063-7761(98)01905-2]

1. INTRODUCTION

It is well known that, as the temperature is reduced, a Peierls transition takes place in quasi-one dimensional conductors, leading to the appearance of a superlattice–a charge density wave or spin density wave, with a period equal to twice the electron Fermi wavelength of the original metal. This transition is accompanied by a partial or complete dielectrization of the electron spectrum in the vicinity of parts of the Fermi surface that are combined during the shift by the wave vector of the charge (spin) density wave. A collective conduction mechanism is associated with the motion of a charge (spin) density wave under the influence of an electric field. Many properties of materials with charge and spin density waves have been studied in detail and these have been reviewed by Grüner.[1] However, a number of questions have not been studied, in particular, the physical mechanisms for conversion of a normal carrier current into a charge (spin) density wave current at a normal metal–Peierls conductor interface.

Until recently, processes taking place inside a charge density wave as it glances between current contacts located on the side surfaces of a ribbon sample of Peierls conductor have been examined. It can now be regarded as reliably established that the current conversion process near contacts of this type is accompanied by a thermally activated slippage of the phase of a charge density wave propagating a substantial distance away from the contact.[2] At the same time, the mechanism by which electrons penetrate from the metal into the charge density wave has not been studied in detail until recently. Brazovskii[3] assumed that penetration of electrons with energies greater than the Peierls gap Δ_P into a charge density wave leads to the formation of solitons, whose interactions with one another cause the formation and growth of dislocations that create a phase slippage between glancing and fixed charge density waves. Here it was assumed that electrons with energies below Δ_P are reflected from the metal–Peierls conductor interface and do not participate in the current conversion.

On the other hand, there is a formal analogy between Peierls conductors and superconductors, since the condensed state in both is described by an ordering parameter $\Delta=|\Delta|\exp(i\varphi)$ whose amplitude determines the energy gap in the spectrum of single-particle excitations, while the derivative of the phase (with respect to the time in Peierls conductors and with respect to position in superconductors) is proportional to the contribution of the condensed electrons to the electric current density. While the condensate in superconductors is formed by pairs of electrons with opposite momenta, a charge density wave can be visualized as a condensate of bound pairs of electrons and holes whose momenta differ by the magnitude of the wave vector of the charge density wave. This similarity leads us to expect an effect at a metal–Peierls conductor interface that is analogous to the Andreev reflection effect in superconductors,[4] reflection accompanied by transformation of a normal carrier current into a charge density wave, at incident electron energies lower than the Peierls gap.

It has been found theoretically[5,6] that, after being reflected, an electron with energy below Δ_P incident on a Peierls conductor from a normal metal moves along the tra-

jectory of the incident particle, i.e., all three components of the velocity vector $\mathbf{v}=d\varepsilon/d\mathbf{p}$ change during reflection. The difference from Andreev reflection is that the sign of the charge of the reflected particle does not change, and this must cause an additional resistance at a metal–Peierls conductor interface.

The first experiments to search for sub-gap reflection at metal–Peierls conductor interfaces were done using van Kempen's method[7] in a study of point contacts formed with an intermediate metal layer sputtered onto the face of a $K_{0.3}MoO_3$ crystal perpendicular to the chain direction.[8] These results are in qualitative agreement with the model[5,6] assuming the existence of a barrier at the metal–Peierls conductor interface analogous to the barrier which arises at a normal metal–superconductor contact.[9] The reflected signal, however, was an order of magnitude lower than predicted.[5,6]

A recent theoretical paper[10] examines the transport properties of metal–Peierls conductor–metal heterostructures. Unlike the earlier work,[5,6] which examined the static reflection of an electron at a barrier owing to the Peierls gap Δ_P, the later work[10] took into account the movement of the condensate from the boundary when an incident electron is reflected. Given that the reflection of a carrier from a metal–superconductor interface is not accompanied by a change in the sign of the charge of the reflected particle, the transformation of the charge in the condensate which occurs during Andreev reflection in superconductors does not take place. However, electron–hole pairs moving from the interface carry away twice the momentum of an electron incident normal to the interface ($\sim 2\hbar k_F$), which corresponds to the momentum of a pair in a charge density wave, i.e., during reflection of this type the momentum transfer is analogous to charge transport during Andreev reflection in superconductors. In other words, a charge–momentum symmetry shows up during reflection from a metal–Peierls conductor interface without charge transport, while during Andreev transport there is no momentum transfer. It is evident that this reflection mechanism (we shall refer to it as the analog of Andreev reflection) can be manifested only when the incident electron moves along the chains, since the component of the momentum along the chains will be less than p_F for oblique incidence. As for the experiment,[8] one consequence of this must be that the reflected signal depends on the area of the point contact.

Another, recently proposed mechanism for the reflection[11] includes the possibility of both mirror reflection of carriers incident on a metal–Peierls conductor interface and reflection at an angle determined by the projection of the wave vector of the charge density wave, **Q**, on a plane perpendicular to the direction of the chains. (In general, **Q** is not parallel to the chains.[1]) By analogy with the coherent scattering of electrons on a crystalline lattice, this type of reflection was referred[11] to as Bragg reflection from the electron crystal. For $\mathbf{Q}=(2k_F,Q_y,0)$ and a wave vector of an electron escaping from an emitter with components (k_x,k_y), reverse trajectories of the carrier into the emitter can develop for $k_y+Q_y=-k_y$, i.e., when the angle to the normal is $\phi=\arcsin(-Q_y/2k_F)$. In the case of mirror reflection, the condition for reverse trajectories, $k_y=-k_y$, is satisfied only for normal incidence, $k_y=0$. In both cases, the reflected signal should also be proportional to the emitter area. It has been shown[11] that the reflection coefficient, as for the previous mechanism,[10] should fall off sharply when the energy of the incident particles exceeds Δ_P.

The purpose of this paper is to clarify which reflection mechanisms operate and under which conditions by varying the experimental configuration (contact area, metal layer thickness, emitter orientation relative to the sample crystal).

2. EXPERIMENTAL TECHNIQUE

Single crystals of ''blue'' bronze $K_{0.3}MoO_3$ with a transverse cross section of 10^3-10^4 μm^2 were used as samples in this work. The Peierls transition temperature, $T_p=181-183$ K, was determined from the temperature dependence of the resistance. The threshold electric field for a glancing charge density wave was determined from the current–voltage characteristic and was 150–500 mV/cm at $T=77$ K. The single crystals were provided by the Center for Low-Temperature Research (CRTBT-CNRS, Grenoble).

To study the processes taking place at a metal–Peierls conductor interface we have used an experimental measurement scheme based on the single-contact method for investigating Andreev reflection in superconductors.[7] The physical basis of this method is illustrated in Fig. 1b for the example of a superconductor. A thin film of normal metal of thickness d is sputtered onto the superconductor and the characteristics of a point contact between a normal metal and this film (a metal–(metal–Peierls conductor) heterostructure) are studied. Electrons are injected through a point contact into a solid angle of 2π and undergo reflection at the metal–superconductor interface. In the case of a ballistic regime, $d \ll l$, and for $a \ll d$ (l is the mean free path and a is the diameter of the point contact), only quasiparticles which undergo Andreev reflection will pass back through the point contact, producing an excess current, which shows up in the current–voltage characteristic of the contact as a reduction in the differential resistance. This method has been used successfully for studying Andreev reflection in traditional superconductors,[12,13] as well as in high-temperature oxide superconductors[14,15]

The geometry of the present experiment is shown in Fig. 1a. The current–voltage characteristics and their first derivatives were measured for point contacts of a normal metal with an Au–$K_{0.3}MoO_3$ heterostructure. Point contacts were formed with crystal planes perpendicular and parallel to the orientation of the chains. A gold coating was deposited by laser sputtering. The layer thicknesses for the different samples were 50 and 100 nm, less than or of order the mean free path in gold at $T=77$ K, which is usually 150–200 nm for gold films. Electrochemically sharpened thin (diameter 30–70 μm) copper and gold wires were used as normal counterelectrodes. The radius of curvature of the needle tips was less than 1μm. A point contact was formed directly at low temperature using a precision mechanical motion transfer system.[16] The characteristics of several point contacts were measured for each sample. The experiments were done at $T=77$ K and at a temperature above T_P for two orienta-

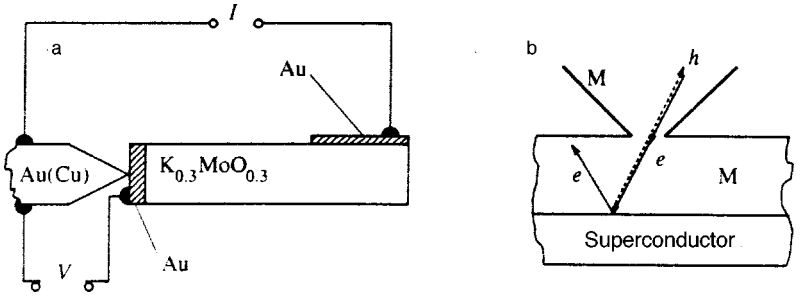

FIG. 1. (a) The experimental geometry and (b) a diagram of the possible reflection processes in a single-contact scheme for measuring Andreev reflection: the continuous lines represent ordinary mirror reflection and the dashed line, reflection of the Andreev type. (M denotes metal.)

tions of the point contact: parallel and perpendicular to the chains.

3. EXPERIMENTAL RESULTS

Nonlinear features associated with the appearance of a Peierls gap were observed in the characteristics of the Au–Au–$K_{0.3}MoO_3$ point contacts only at temperatures below the Peierls transition temperature. Above $T_P = 183$ K, the current–voltage characteristics of the point contacts are symmetric and are ohmic in character with a small rise in the resistance proportional to the square of the voltage. No other features were observed at these temperatures for any orientations of the point contacts. The results of the measurements for $T < T_P$ are presented below.

3.1. Orientation of the point contact along the chains

The physical picture of reflection at a metal–Peierls conductor interface may be clarified somewhat by experimental data on the evolution of the characteristics of the Au–Au–$K_{0.3}MoO_3$ point contacts as the diameter of the point contact and the thickness of the metal layer are varied. In fact, since the Andreev-type reflection predicted in Refs. 5 and 6 must occur for all injection directions through the point contact (we assume that the carrier injection is isotropic within a solid angle of 2π), in the ballistic regime the experimental variation in the reflected signal amplitude should not depend on the diameter d of the point contact. On the other hand, when there is a large contribution from normal reflection, as predicted in Refs. 10 and 11, the signal amplitude should increase as the diameter of the point contact is increased (i.e., as its resistance is lowered), since the fraction of normally reflected particles then increases as $(a/d)^2$. For a clear manifestation of the effect, the effect of scattering has to be minimized, and this can be done by reducing the thickness of the normal layer.

Figure 2 shows typical plots of the differential resistance $R_d = dV/dI$ as a function of the voltage V on Au–Au–$K_{0.3}MoO_3$ point contacts at $T = 77$ K for two samples. The nonlinearity of the $R_d(V)$ curves is caused by a combination of two effects: Joule heating and the reflection of injected carriers from the metal–Peierls conductor interface. The latter effect typically[8] has a minimum in R_d at $V = 0$ and maxima at $eV_0 = \pm\Delta_P$. It is clear from the figure that these curves are analogous to the curves given in Ref. 8, but the signal amplitude is greater, while the peaks in the excess resistance at $V = \pm V_0$ corresponding to the Peierls energy gap are much more distinct for the sample with the thinner metal layer, which may be a result either of reduced scattering or of the change in the ratio a/d.

The substantial asymmetry of the curves is most likely caused by a distortion in the chemical potential near the metal–Peierls conductor interface, as well as at the metal–semiconductor contact. (See Fig. 5 of Ref. 8.) A final conclusion regarding this question, however, will require additional data, specifically, on the characteristics of metal–Peierls conductor point contacts.

For most of the $R_d(V)$ curves at high bias voltages V, the curve emerges into a square law dependence for the increment in resistance corresponding to Joule heating of the contact. Typically the heating is by 1–3 K for the least resistive contacts and, given that the linear size of the heat release region is large compared to the region over which the voltage drops,[17] the heating does not affect the measurement results significantly. In this region, the differential resistance is determined by the resistance of the metal–metal contact $R_{dN}(V)$, as such. Assuming that the increment in the resistance behaves as $\Delta R_{dN}(V) \propto V^2$ (dashed curve in Fig. 2), we can isolate the excess resistance $R_d - R_{dN}$ associated with the presence of a Peierls gap in the Peierls conductor and follow the change in the reflected signal amplitude as the contact area is varied. Figure 3 shows the amplitude of the normalized differential resistance $A^* = A - 1$, where we have written $A = R_d/R_{dN}$, as a function of the voltage V for sample No. 4 for contacts with different values of $R_{dN}(0)$. As can be

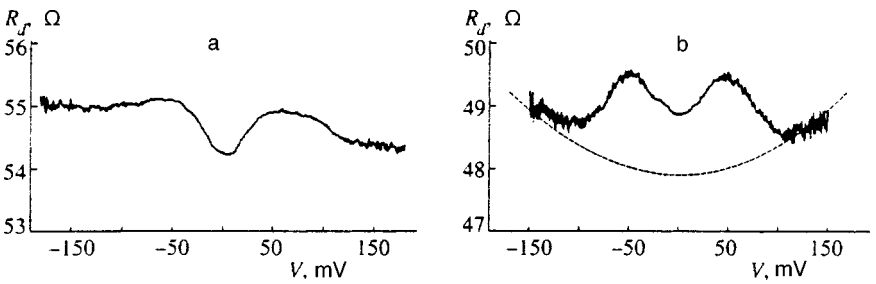

FIG. 2. The differential resistance $R_d = dV/dI$ as a function of voltage V for Au–Au–$K_{0.3}MoO_3$ point contacts at $T = 77$ K: (a) sample No. 1, $d = 100$ nm, $R_d(0) = 54.26\ \Omega$; (b) sample No. 4, $d = 50$ nm, $R_d(0) = 48.87\ \Omega$. The dashed curve shows the voltage dependence of the resistance for the intrinsic metal–metal contact, $R_{dN}(V)$.

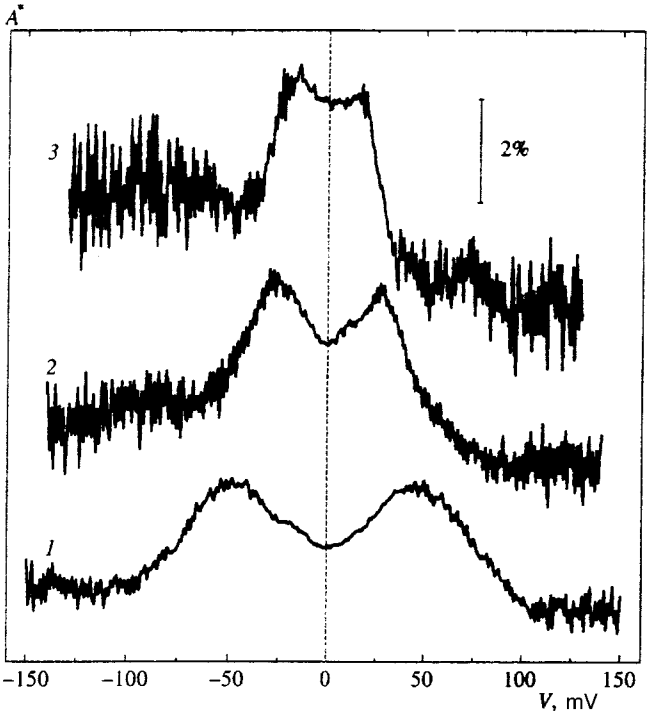

FIG. 3. The amplitude of the normalized differential resistance $A^* = R_d/R_{dN} - 1$ as a function of voltage V for sample No. 4 at point contacts with $R_{dN}(0) = 48.3\ \Omega$ (*1*), 26.7 Ω (*2*), and 15.8 Ω (*3*) at $T = 77$ K. The point contacts are oriented along the chains.

seen from the figure, $A^*(0)$ increases as the contact resistance is raised. For $R_d(0) > 50\ \Omega$, the Peierls gap, defined as the separation $2eV_0$ between the peaks in the differential resistance, did not vary, being equal to $2\Delta_P \simeq 100$ meV (curve *1*). One unusual effect, however, should be noticed: there is a large reduction in the gap as the contact resistance is lowered for contact resistances $R_d(0) < 50\ \Omega$.

3.2. Orientation of the point contact perpendicular to the chains

Measurements similar to those described above were taken at $T = 77$ K with the orientation of the point contact perpendicular to the chains in the $[\bar{2}01]$ direction. Figure 4 shows the differential resistance R_d as a function of the voltage V for sample No. 5 ($d = 50$ nm) with two different resistances $R_{dN}(0)$. The figure shows that these curves are very similar to the ones described above. There are, however, two fundamental differences. First, the position of the peaks at $\pm V_0$ in the $R_d(V)$ curves did not change for different thicknesses of the metal layer and for all point contact resistances and corresponded to a gap of $2V_0 = 2\Delta_p \simeq 100$ mV. Second, the reflected signal amplitude is substantially lower ($< 1\%$) than when the point contact is oriented along the chains for the same metal layer thickness; this made it impossible to determine correctly the effect of changes in the point contact diameter on the reflected signal.

4. DISCUSSION OF RESULTS

We begin the discussion with the experiments on charge injection along the chains, i.e., along the b axis. As can be seen from Fig. 3, the $A(V) = A^* + 1$ curve typically has a minimum at $V = 0$ and maxima at $|eV_0| = \Delta_P$. This behavior can be explained by the presence of carrier reflection, in principle, caused by any of the mechanisms considered[5,6,10,11] with a barrier associated with the nonideal character of the interface, by analogy with the Tinkham–Blonder–Klapwijk model[9] developed for a metal-superconductor interface. The fundamental point is that the $A(V)$ curve is a mirror reflection of the corresponding curve for a metal–superconductor interface relative to the line $A = 1$, i.e., an excess resistance is observed for $|eV_0| < \Delta_P$. This indicates that the reflected particles do not change their charge state; for example, electrons are reflected as electrons. The role of the barrier, as in the case of a metal–superconductor interface,[9] reduces to lowering the probability of penetration of the particles into the condensate and, therefore, to suppressing interactions of the incident electrons with the condensate at low voltages. Without a barrier and for $T = 0$, the $A(V)$ curve should have the step function form

$$A(V) = \begin{cases} \text{const} > 1 & |V| < \Delta_P/e, \\ 1, & |V| > \Delta_P/e. \end{cases}$$

A drop in $A(V)$ for $|V| > \Delta_P/e$ is implied by all the models considered in the Introduction.[5,6,10,11] The difference lies in the scale of the effect. According to Refs. 5 and 6, in the ballistic regime all the trajectories with a length shorter than the mean free path are reversed, so that the scale of the amplitude A^* of the excess resistance is $A^* \sim 1$. For mechanisms analogous to Andreev reflection, i.e., mirror or Bragg reflection, only normal trajectories (the analog of Andreev reflection, mirror reflection) or the trajectories at the Bragg angle (Bragg reflection) are reversed, so that the amplitude A^* should be small, on the scale of a^2/d^2; that is, A^* must be much smaller than unity, since in the experiments a was 10–20 Å and $d \sim 500$ Å. This estimate for d was obtained from the Sharvin formula[18] $R \sim \rho l/a^2$, where ρ is the resistivity and l is the mean free path. As Fig. 3 implies, the

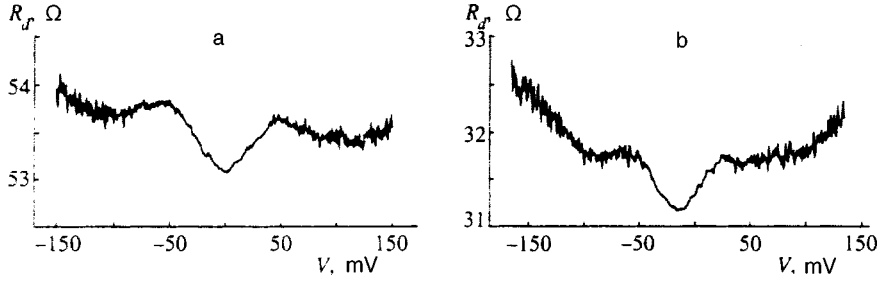

FIG. 4. The differential resistance $R_d = dV/dI$ as a function of voltage V for Au–Au–$K_{0.3}MoO_3$ point contacts at $T = 77$ K, sample No. 5 with $R_{dN}(0) = 53.08\ \Omega$ (a) and 31.16 Ω (b). The point contacts are oriented perpendicular to the chains.

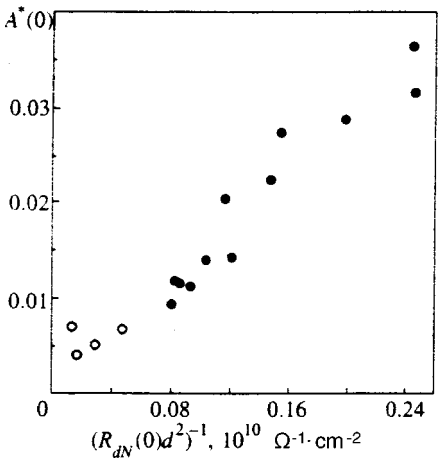

FIG. 5. The amplitude $A^*(0)$ at $T=77$ K as a function of the parameter $(R_{dN}(0)d^2)^{-1}$ (● $d=50$ nm, ○ $d=100$ nm). The point contacts are oriented along the chains.

magnitude of this effect is quite small ($\sim 1-3\%$), which evidently allows us to neglect mechanisms of the type proposed in Refs. 5 and 6.

It should be noted that the amplitude A^* increases as the film thickness d is reduced (Fig. 2). This may be related to the reduced influence of scattering for smaller d, or be the result of a dependence of the form a^2/d^2. To verify this dependence, experiments with different contact diameters (different resistances) were set up. Some of the results are shown in Fig. 3. We note the following:

(1) As the degree of clamping is increased (the contact resistance is reduced), the picture does not change qualitatively for $R_d(0)>50\ \Omega$. The positions of the peaks at $\pm V_0$ is unchanged and corresponds to a Peierls gap of $2V_0 = 2\Delta_P = 100$ meV, in agreement with optical measurements and measurements of the thermally activated conductivity along the b axis.[19]

(2) For $R_d(0)<50\ \Omega$ the value of $2\Delta_P$ begins to decrease. We attribute this distortion to nonequilibrium effects caused by strong current injection into the Peierls conductor. (For more detail, see below.)

(3) The reflected signal at $V=0$ (under conditions of minimal injection) is proportional to a^2/d^2 (Fig. 5). Behavior of this sort is implied by two of the models, the analog of Andreev reflection[10] and Bragg and mirror reflection,[11] but contradicts the model proposed in Refs. 5 and 6.

In analyzing the possible contribution of Bragg scattering, it is necessary to keep in mind that the wave vector of a charge density wave in $K_{0.3}MoO_3$ has the form[20]

$$\mathbf{q}=0\mathbf{a}^*+q_b\mathbf{b}^*+0.5\mathbf{c}^*$$

(the corresponding lattice parameters are $a=16.23$ Å, $b=7.55$ Å, and $c=9.86$ Å; $\beta=94.89°$), where q_b is the projection of $\mathbf{q}$ on the b axis, with $q_b\sim 2k_F$. The condition for a reversed trajectory during Bragg reflection gives $q_c=-2k_c$ (see the Introduction), where k_c is the projection of the wave vector of the incident electron on the c axis. For $k_c<k_F$ this means that q_c must be less than $2k_F$ and, therefore, less than q_b, i.e., the condition $q_c<q_b$ must be satisfied. An approximate estimate gives $q_c\simeq 0.5/c\simeq 5\cdot 10^6$ cm^{-1} and $q_b \simeq 0.25/b\simeq 3\cdot 10^6$ cm^{-1}, and we see that Bragg reflection does not produce reversed trajectories in $K_{0.3}MoO_3$ because of the very large projection q_c.

This analysis, therefore, leaves the possibility of two mechanisms for reflection of normal carriers during injection along the chains, the analog of Andreev reflection or mirror reflection, but it is not possible to separate the contributions from these two mechanisms using only longitudinal measurements. The picture can be clarified by comparing data obtained in two injection directions: along the chains (parallel to the b axis) and perpendicular to the chains (perpendicular to the b axis in the $[\bar{2}01]$ direction). In analyzing the results for transverse injection, we begin by noting two important points:

First, the $R_{d\perp}(V)$ curves have the same shape as the $R_{d\parallel}(V)$ curves for large R_d. Here the positions of the peaks in $R_d(V)$ are the same (± 50 mV) as for longitudinal injection. This indicates that the Peierls gap is isotropic parallel and transverse to the chains in the plane $(\mathbf{b},2\mathbf{a}^*-\mathbf{c})$. Until recently, a result of this sort had been obtained only from indirect, averaged measurements of thermally activated conduction along these directions.[19]

Second, during injection perpendicular to the chains there can be no contribution from the mechanism analogous to Andreev reflection,[10] since then the longitudinal component of the momentum (along the chains) is essentially zero or at least much smaller than p_F, i.e., momentum cannot be transferred to the moving condensate during reflection. Thus, the observed response during this kind of injection can only be attributed to the Bragg or mirror reflection mechanism. Then, as noted previously,[11] the contribution from mirror reflection should dominate because of the large difference between the effective masses of the carriers in the metal and in the Peierls conductor (in a direction perpendicular to the chains).

On the other hand, the contribution of mirror reflection during injection perpendicular to the chains should be no less than during injection along the chains (because of the lower effective mass along the chains). However, in the experiment it was found that the signal amplitude during injection along the chains is systematically roughly a factor of three higher when the other conditions (film thickness, contact resistance, etc.) are the same. (See Figs 3 and 4, for example.) This indicates the presence of an additional contribution beyond mirror reflection during longitudinal injection, which we attribute to a mechanism analogous to Andreev reflection. Besides the estimates given above, this hypothesis is supported, first, by a correct orientation for the appearance of this mechanism, second, by a linear dependence of this contribution (as of the entire signal) on the parameter a^2/d^2, and third, by the $\Delta_P(I)$ curve at high currents, which indicates that a charge density wave moves in the carrier injection region (see below).

Analysis of data for high current injection

We now return to examining the results for injection along the chains at low resistance contacts (Fig. 4). The figure implies that as the contact resistance is lowered, there is a noticeable reduction in the Peierls gap, which we define as

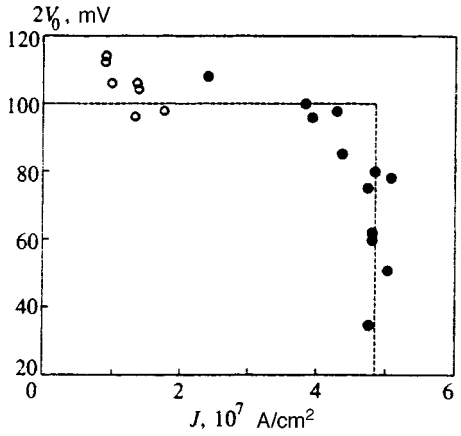

FIG. 6. The voltage $2V_0$ as a function of the current density J through a metal-Peierls conductor interface with normal injection of quasiparticles (● $d=50$ nm, ○ $d=100$ nm) at $T=77$ K. The point contacts are oriented along the chains.

the distance between the peaks $R_d(\pm V_0)$, and the reduction in the gap set in only for contact resistances $R_{d0}(0) < 50\ \Omega$. For $R_d > 50\ \Omega$ the gap $2\Delta_P \simeq 100$ meV did not vary, which fact is, as noted above, in good agreement with data from optical and thermal activation measurements of the conductivity.

Lowering the resistance leads to an increase in the current through the contacts required to attain voltages corresponding to the gap features. Thus, it is logical to assume that one possible reason for the reduction in Δ_P as the contact resistance is varied may be suppression of the gap owing to a higher current density through the metal–Peierls conductor interface. In our case, $a \ll d$ for all the contacts, where d is the thickness of the gold film. Given that injection is isotropic through a point source, we can estimate the current density through the metal–Peierls conductor interface as

$$J = \frac{I \cos^2 \phi}{2\pi d^2},$$

where I is the total current through the contact and ϕ is the injection angle. Figure 6 shows a plot of the voltage $2V_0$ (it will be shown below that this is not equal to the equilibrium value of $2\Delta_P/e$) as a function of the parameter I/d^2, which is proportional to the current density through the interface, for samples with normal layer thicknesses $d=50$ and 100 nm. This dependence has a distinct threshold character, i.e., V_0 is essentially independent of J up a critical value of the current density through the metal–Peierls conductor interface, $J_c = 4.8 \cdot 10^7$ A/cm^2, at which there is a sharp drop in V_0; this corresponds to suppression of the Peierls gap. One can say that Δ_P is essentially independent of J up to J_c, after which the gap goes discontinuously to zero. Evidently, in this interpretation, for small contact resistances the voltage V_0 corresponds to a nonequilibrium value of the gap (less than the equilibrium value Δ_P), since it is measured at current densities close to the critical value corresponding to the maximum propagation speed of a charge density wave. It should be specially noted that this interpretation is possible only in the case where the injected carriers interact with the charge density wave so that momentum is transferred to electron–hole pairs in the condensate. In fact, an estimate of the propagation speed v of the charge density wave corresponding to the current density J_c for normal injection using the simplest formula $v = J_c/ne$, where n is the density of condensed carriers, yields $v = 0.6 \cdot 10^5$ cm/s, which is close to the sound speed in this material. But the propagation speed of the charge density wave obviously cannot exceed the sound speed.

In principle, one can recover the shape of the reflection spectra without current suppression of the gap. As can be seen from Fig. 3, the depth of the minimum at $V=0$, which is determined by the potential barrier at the metal–Peierls conductor interface, changes substantially when the contact resistance is reduced. At the same time, there are no reasons for a large spatial variation in the barrier at the sample surfaces (and, therefore, from contact to contact). Then the reflection spectrum for curve *3* of Fig. 3, $A_3^*(V)$, without current suppression is obtained by multiplying the amplitude of the excess resistance $A_1^*(V)$ for curve 1 by the normalizing factor $\eta = A_3^*(0)/A_1^*(0)$. Following the same procedure with curve 2, we obtain the spectra shown in Fig. 7, which show clearly that the reflected amplitude would increase noticeably without current suppression, even for $V=V_0$.

5. CONCLUSIONS

(1) The contribution of reflected quasiparticles to the contact resistance is sharply reduced when the energy of the incident particles exceeds the Peierls gap of the charge density wave, which means that the injected carriers interact with the condensate of the charge density wave.

(2) When normal carriers are reflected from a metal–Peierls conductor interface, their charge state does not change, unlike in Andreev reflection at a metal–superconductor interface.

(3) It has been shown that the barrier effect owing to the nonidealness of a metal–Peierls conductor interface shows up as a suppression of the gap feature near zero bias voltage similar to that observed at a metal–superconductor interface.

(4) Local measurements show that in $K_{0.3}MoO_3$ the gap is isotropic in the [010] direction along the chains and in the $[\bar{2}01]$ direction perpendicular to the chains.

(5) The existence of an excess resistance during injection perpendicular to the chains in the $[\bar{2}01]$ direction is evidence of mirror reflection from a barrier of height Δ_P.

(6) A comparison of the measurement data for injection along and perpendicular to the chains shows that during longitudinal injection there is reflection involving momentum transfer from the incident particles to the condensate of moving pairs in the charge density wave (an analog of Andreev reflection).

(7) The experimental results indicate that the Peierls gap is suppressed by the current for high injection levels along the chains corresponding to a current density $\simeq 5 \cdot 10^7$ A/cm^2.

We thank S. N. Artemenko, S. N. Zaĭtsev-Zotov, V. Ya. Pokrovskiĭ, P. Monso, and R. Escudeiro for useful discus-

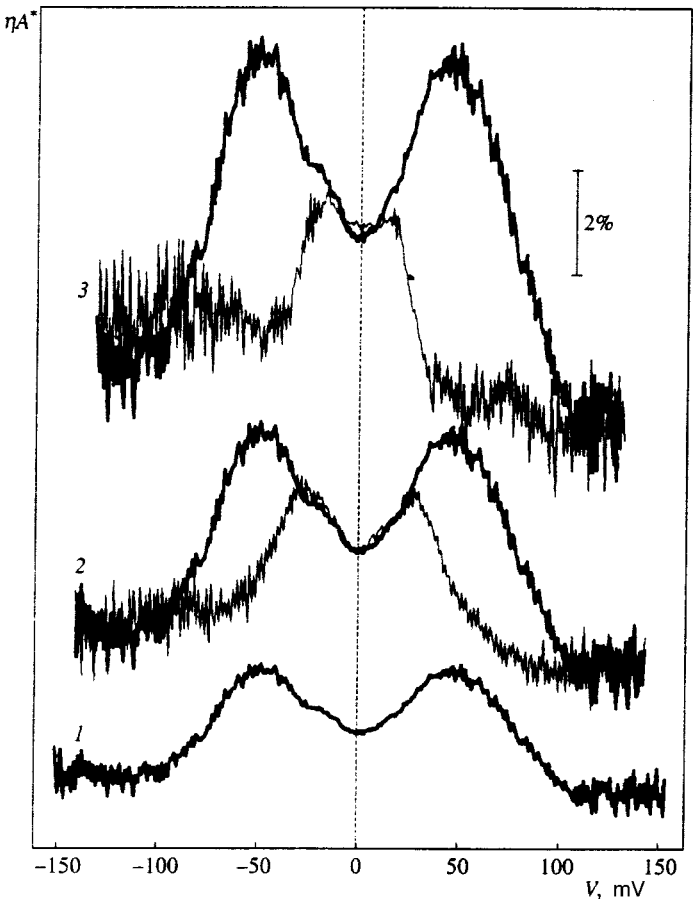

FIG. 7. The amplitude $\eta A^*(V)$ for sample No. 4 without current suppression of the Peierls gap for contacts with $R_{dN}(0)=48.3\ \Omega$ (*1*), 26.7 Ω (*2*), and 15.8 Ω (*3*) at $T=77$ K. The point contacts are oriented along the chains.

sions of the experimental results. This work was supported by the Russian Fund for Fundamental Research through Grants Nos. 95-02-04456 and 97-02-17108.

*)E-mail: gorl@mail.cplire.ru

[1] G. Grüner, Ed., *Density Waves in Solids*, Addison-Wesley, Reading, MA (1994).
[2] T. L. Adelman, M. C. de Lind van Wijngaarden, S. V. Zaitsev-Zotov *et al.*, Phys. Rev. B **53**, 1833 (1996).
[3] S. Brazovskii, in *Charge Density Waves in Solids*, L. P. Gor'kov and G. Grüner, Eds., *Modern Problems in Condensed Matter Sciences*, Vol. 25, North Holland, Amsterdam (1989), p. 425.
[4] A. F. Andreev, Zh. Éksp. Teor. Fiz. **46**, 1823 (1964) [Sov. Phys. JETP **19**, 1228 (1964)].
[5] A. L. Kasatkin and É. A. Pashitskiĭ, Fiz. Nizk. Temp. **10**, 1222 (1984) [Sov. J. Low Temp. Phys. **10**, 640 (1984)].
[6] A. L. Kasatkin and É. A. Pashitskiĭ, Fiz. Tverd. Tela **27**, 2417 (1985) [Sov. Phys. Solid State **27**, 1448 (1985)].
[7] P. A. M. Benistant, A. P. van Gelder, and H. van Kempen, Phys. Rev. B **32**, 3351 (1985).
[8] A. A. Sinchenko, Yu. I. Latyshev, S. G. Zybtsev *et al.*, JETP Lett. **64**, 285 (1996).
[9] G. E. Blonder, M. Tinkham, and T. M. Kkapwijk, Phys. Rev. B **25**, 4515 (1982).
[10] B. Rejaei and G. E. W. Bauer, Phys. Rev. B **54**, 8487 (1996).
[11] S. N. Artemenko and S. V. Remizov, JETP Lett. **65**, 53 (1997).
[12] P. C. van Son, H. van Kempen, and P. Wyder, Phys. Rev. Lett. **59**, 2226 (1987).
[13] H. F. C. Hoevers, M. G. D. van der Grinten, P. L. H. Jennen *et al.*, J. Phys.: Condens. Matter **6**, 65 (1994).
[14] T. W. Jing, Z. Z. Wang, and N. P. Ong, Physica C **162–164**, 1061 (1989).
[15] V. F. Elesin, A. A. Sinchenko, A. A. Ivanov, and S. G. Galkin, Physica C **213**, 490 (1993).
[16] A. A. Sinchenko, Fiz. Nizk. Temp. **15**, 438 (1989) [Sov. J. Low Temp. Phys. **15**, 247 (1989)].
[17] I. O. Kulik, A. N. Omel'yanchuk, and R. I. Shekhter, Fiz. Nizk. Temp. **3**, 1543 (1977) [Sov. J. Low Temp. Phys. **3**, 740 (1977)].
[18] Yu. V. Sharvin, Zh. Éksp. Teor. Fiz. **48**, 984 (1965) [Sov. Phys. JETP **21**, 655 (1965)].
[19] C. Schlenker, J. Dumas, C. Escribe-Filippini *et al.*, Philos. Mag. B **52**, 643 (1985).
[20] M. Chedira, J. Chenavas, M. Marezio, and J. Marcus, J. Solid State Chem. **57**, 300 (1985).

Translated by D. H. McNeill

Paramagnetic labeling as a method for the soft spectroscopy of electronic states

M. N. Kiselev and A. S. Mishchenko*)

Kurchatov Institute Russian Scientific Center, 123182 Moscow, Russia
(Submitted 28 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1843–1865 (May 1998)

A self-consistent microscopic theory of the relaxation of the crystal-field levels of an impurity ion in a state with an integer valence implanted in a normal metal is devised. A microscopic approach based on the Coqblin–Schrieffer–Cooper approach, rather than the formal model of the sf exchange interaction, makes it possible to take into account the specific details of both the crystal-field states of the impurity ion and the electronic band spectrum of the metal. A new method for the soft spectroscopy of electronic states based on measurements of the temperature dependence of the width $\Gamma_{MM'}(T)$ of transitions between the crystal-field states $|M\rangle$ of a paramagnetic ion implanted in the compound being studied is proposed. To make specific use of this method in neutron and optical spectroscopy, a classification of the types of temperature dependence of the natural relaxation width $\gamma_M(T)$ of the levels is devised, and procedures for possible experimental methods are proposed. A nonzero value of the natural relaxation width $\gamma_G(T)$ of the crystal-field ground state $|G\rangle$ of an impurity ion at zero temperature is obtained within the proposed self-consistent model, but is beyond the scope of perturbation theory. It is shown that the widely accepted estimate of the characteristic temperature of Kondo systems $T^*=\Gamma_G(T=0)/2$ from the quasielastic scattering width at zero temperature $\Gamma_G(T=0)/2$ is incorrect in the case of strong relaxation in a system with soft crystal fields. The proposed model is applied to the quantitative analysis of the relaxation of the crystal-field levels of paramagnetic Pr^{3+} ions implanted in $CeAl_3$ and $LaAl_3$. The results of the calculations are in quantitative agreement with the experimental data. © *1998 American Institute of Physics.* [S1063-7761(98)02005-8]

1. INTRODUCTION

The methods that have been developed for studying electronic states in metals (angle-resolved photoemission spectroscopy;[1] quantum oscillations of the magnetic susceptibility,[2] conductivity,[3] magnetostriction,[4] and elastic moduli[5] associated with the de Haas–van Alphen effect; infrared spectroscopy;[6] Raman scattering;[7] etc.) provide complementary information regarding the structure of electron spectra. A comparison of the experimental data obtained by different methods with the results of band calculations of the electronic structure provides fairly reliable data on the properties of the compounds studied.

The methods for investigating electronic states can be divided into ''hard'' and ''soft'' methods. In the case of hard spectroscopy, the influence of the measurement process on the system exceeds the scales W^* of the characteristic interactions forming the electronic spectrum of the system (in Kondo systems W^* is of the order of the Kondo temperature T_K; in variable-valence systems W^* is of the order of the valence fluctuations). Therefore, compounds with strong electron correlations, which have low-energy modes in the spectrum of elementary excitations, can be investigated most effectively by soft spectroscopic methods, in which the measurement process does not destroy the eigenstates of the system being investigated. The conditions imposed on spectroscopic measurements by the softness of the elementary excitations in variable-valence and Kondo systems greatly restricts the set of methods that are applicable to the investigation of highly correlated systems. For example, the interpretation of photoemission measurements (because of the large energy transfers in the measurement process) and data from methods based on de Haas–van Alphen oscillations (because of the large magnetic fields, which can destroy the structure of soft excitations) requires a special investigation of the influence of the measurement process on the low-energy properties of the compound being studied. Therefore, the development of new soft spectroscopic methods for highly correlated electronic systems is an important undertaking.

This paper proposes a method for analyzing the electronic structure based on measurements of the temperature dependence of the relaxation of crystal-field levels of an impurity ion which has special properties (a paramagnetic label) and is implanted in the compound being investigated. A similar idea for investigating semiconductor compounds by an electron paramagnetic resonance technique was proposed back in Ref. 8. The method discussed in this paper relies on the technique of measuring the neutron or optical response of the system and is intended for studying metallic compounds. A spectroscopic procedure employing a paramagnetic label can be divided into two stages. In the first stage highly complete information on the energies and wave functions of the paramagnetic label P must be obtained. To this end a combined study (neutron scattering or Raman scattering measurements; magnetic susceptibility and specific heat mea-

surements) must be made of a reference single crystal of $P_l\{B\}$ ($\{B\}$ is the chemical formula without the paramagnetic label). In the second stage small quantities of the A ions in the compound $A_l\{B\}$ under investigation are replaced by the paramagnetic label P. Scrutiny of the temperature dependence of the relaxation of crystal-field levels of the paramagnetic label P in the compound $(A_{1-x}P_x)_l\{B\}$ can provide unique information regarding the electronic structure of the compound under investigation when several conditions are fulfilled. First, the inequality $x \ll 1$ is a necessary condition, which allows us to treat the relaxation of the crystal field at the paramagnetic label as a purely single-ion effect. Second, it must be shown that the structure of the crystal field of the paramagnetic label P in $(A_{1-x}P_x)_l\{B\}$ does not differ significantly from the structure of the crystal field in the pure reference crystal of $P_l\{B\}$. Fulfillment of the second condition has already been demonstrated for several compounds, in which the main contribution to the formation of the crystal field is made by the nearest neighbors from a formula unit of $\{B\}$, and hence the structures of the crystal fields of the paramagnetic label P in $P_l\{B\}$ and $(A_{1-x}P_x)_l\{B\}$ are practically identical. Examples of such compounds include RAl_3,[9–11] RNi_5,[12–14] and RNi[14–16] (R is a rare-earth ion).

It should be noted that the existing methods for calculating the temperature dependence of the relaxation of crystal-field states cannot be applied to the analysis of specific, highly correlated systems. Some of the methods employ the formal Hamiltonian, i.e., one which is not related in any way to the electronic structure, of the sf model.[17–19] Another deficiency of the previously developed methods is the use of nonself-consistent second-order perturbation theory,[17,18,20,21] which is inapplicable in the case of the large relaxation widths characteristic of highly correlated systems.

The goal of the present work is to devise a self-consistent theory for the relaxation of crystal-field levels, which can serve as a tool for studying the electronic structure of particular, highly correlated electronic systems with strong relaxation broadening. Section 2 presents the derivation of a microscopic interaction Hamiltonian, an analysis of the differences between it and the formal Hamiltonian of the sf model, and a discussion of the Coqblin–Schrieffer model. In Sec. 3 self-consistent equations are obtained for the natural relaxation widths of the crystal-field levels, and their influence on the cross section for magnetic inelastic neutron scattering is analyzed. In Sec. 4 qualitatively different types of temperature dependence of the relaxation width are classified. The effects associated with departure from the weak-relaxation approximation are analyzed in Sec. 5. In Sec. 6 the conditions which must be satisfied by the paramagnetic label are analyzed in detail, and experimental procedures which provide the most easily interpreted information are presented. The results of measurements of the relaxation widths of the crystal-field states of the paramagnetic label Pr^{3+} in the compounds $LaAl_3$ and $CeAl_3$ are considered in Sec. 7. The conclusions are presented in Sec. 8.

2. SPECIFIC DETAILS OF THE INTERACTION OF CRYSTAL-FIELD STATES WITH CONDUCTION ELECTRONS IN THE COQBLIN–SCHRIEFFER MODEL

The interpretation of the relaxation of a real paramagnetic label in a particular compound requires the formulation of a problem which takes into account both the specific details of the state of the impurity and the features of the electronic structure of the metal. Therefore, the sf exchange Hamiltonian, which is often employed to analyze the relaxation of crystal-field levels,[17–19]

$$\mathscr{H}_{sf}=\sum_{MM'}(f_M^{\dagger}\hat{\mathbf{J}}_{MM'}f_{M'})(c_{\alpha}^{\dagger}\boldsymbol{\sigma}_{\alpha\beta}c_{\beta}) \tag{1}$$

(where M and M' are the indices of the crystal-field states, α and β are the spin indices of the conduction electrons, $\hat{\mathbf{J}}$ is the total momentum operator, and $\boldsymbol{\sigma}$ denotes a Pauli matrix) is unsuitable for analyzing relaxation in a particular system, since it is a purely formal object, which is not related in any way to the features of the electronic structure of the metal or to the real character of the interaction of an impurity with conduction electrons.

The specific features of the relaxation occurring as a consequence of the interaction of an impurity with conduction electrons can be taken into account in the approaches[22–25] based on the Schrieffer–Wolff and Cornut–Coqblin formalisms.[26–28] A scheme permitting a first-principles calculation of the relaxation of a paramagnetic label can be devised within the method proposed in Refs. 22–25. The Anderson Hamiltonian describing an impurity ion[1)] with one f electron implanted in a metal is represented in the form of the sum

$$\mathscr{H}=\mathscr{H}_0+\mathscr{H}_1. \tag{2}$$

Here the first term

$$\mathscr{H}_0=\sum_{\theta\mathbf{k}\sigma}\epsilon_{\theta\mathbf{k}}c_{\theta\mathbf{k}\sigma}^{\dagger}c_{\theta\mathbf{k}\sigma}+\sum_M E_M f_M^{\dagger}f_M + \frac{U}{2}\sum_{MM'}^{M\neq M'} f_M^{\dagger}f_M f_{M'}^{\dagger}f_{M'} \tag{3}$$

describes the subsystem of delocalized conduction electrons with consideration of the single-particle potential of the f subshell (which is treated as a core state) and the subsystem of the crystal field of the f subshell in the single-particle potential created by the conduction electrons. The operator $c_{\theta\mathbf{k}\sigma}^{\dagger}$ ($c_{\theta\mathbf{k}\sigma}$) describes the creation (annihilation) of a conduction electron with the energy $\epsilon_{\theta\mathbf{k}}$, whose state is characterized by the Bloch wave

$$|\theta\mathbf{k}\sigma\rangle=u_{\theta\mathbf{k}}(\mathbf{r})e^{i\mathbf{kr}}|\sigma\rangle \tag{4}$$

with the wave vector $\mathbf{k}$, the band index θ, and the spin projection σ. The operator $f_M^{\dagger}$ (f_M) describes the creation (annihilation) of the crystal-field state $|M\rangle$ with the energy E_M. The wave functions $|M\rangle$ of the states of an f electron transform in accordance with the irreducible representation Υ_M of the point group of the site of the rare-earth impurity ion $\mathscr{G}_{\text{imp}}$:

$$|M\rangle = \sum_{m=-J_{\rm imp}}^{J_{\rm imp}} \Lambda^{M}_{m,J_{\rm imp}} |m\rangle. \tag{5}$$

Here the $|m\rangle$ are spherical harmonics, which describe the projections m of the total angular momentum of the impurity $J_{\rm imp}$, and U is the on-site Coulomb repulsion constant.

For a microscopic calculation procedure we must represent the many-particle interaction of the localized and delocalized subsystems in terms of the nomenclature for the band states of conduction electrons, rather than in the approximation of symmetrized partial waves.[27,28] In this nomenclature the interaction Hamiltonian

$$\mathscr{H}_1 = \sum_{\theta \mathbf{k}\sigma M} V^{M}_{\theta \mathbf{k}\sigma} f^{\dagger}_{M} c_{\theta \mathbf{k}\sigma} + \text{H.c.} \tag{6}$$

describes the mixing of the localized state $|M\rangle$ with the Bloch wave $|\theta \mathbf{k}\sigma\rangle$, and the hybridization parameter

$$V^{M}_{\theta \mathbf{k}\sigma} = \langle \theta \mathbf{k}\sigma | V_{\rm mix}(\mathbf{r}) | M\rangle \tag{7}$$

can be calculated by a band-calculation procedure. In the case of an impurity state with a nearly integer valence (the hybridization scale $|V^{M}_{\theta \mathbf{k}\sigma}|$ is considerably smaller than the distance from the E_M and $E_M + U$ levels to the Fermi energy ϵ_F), the Coqblin–Schrieffer transformation,[26,27] which eliminates the first order with respect to the hybridization from the Hamiltonian, is applicable. As a result, the interaction of the localized and delocalized subsystems is described by elastic and inelastic scattering processes of the conduction electrons on localized crystal-field states of the impurity:

$$\mathscr{H}_{\rm ex} = \sum_{MM'} \sum_{\theta \mathbf{k}\sigma} \sum_{\theta' \mathbf{k}'\sigma'} J^{MM'}_{\theta \mathbf{k}\sigma, \theta' \mathbf{k}'\sigma'} f^{\dagger}_{M} f_{M'} c^{\dagger}_{\theta \mathbf{k}\sigma} c_{\theta' \mathbf{k}'\sigma'}. \tag{8}$$

The interaction constants of the effective Hamiltonian are expressed in terms of quantities which can be determined by band-calculation methods:[22–25]

$$J^{MM'}_{\theta \mathbf{k}\sigma, \theta' \mathbf{k}'\sigma'} = \frac{V^{M'}_{\theta \mathbf{k}\sigma} (V^{M}_{\theta' \mathbf{k}'\sigma'})^{*}}{2} \left[\frac{1}{\epsilon_{\theta \mathbf{k}} - E_M} + \frac{1}{\epsilon_{\theta' \mathbf{k}'} - E_{M'}} \right]. \tag{9}$$

Although the nomenclature of the band states of conduction electrons is adequate in cases where the problem is a first-principles calculation of the parameters, the nomenclature of symmetrized partial waves, which permits the use of symmetry arguments, is more convenient for qualitative analysis. As a result of the standard transformation into the representations $|\theta k M\rangle$ of the partial waves[27,28]

$$c^{\dagger}_{\theta \mathbf{k}\sigma} = \sum_{kM''} \langle \theta \mathbf{k}\sigma | \theta k M''\rangle c^{\dagger}_{\theta k M''} \tag{10}$$

(here $c^{\dagger}_{\theta k M''}$ is the annihilation operator of a conduction electron in the state centered on the impurity ion with the wave number k, the total angular momentum $J_{\rm imp}$, and the angular dependence described by the irreducible representation $\Upsilon_{M''}$), the Hamiltonian of the exchange interaction can be represented in the form

$$\mathscr{H}_{\rm ex} = \sum_{MM'} f^{\dagger}_{M} f_{M'} \sum_{M''M'''} \sum_{kk'} \sum_{\theta\theta'} \Theta^{\theta' k' M'''}_{\theta k M''}(M, M') \times c^{\dagger}_{\theta k M''} c_{\theta' k' M'''}, \tag{11}$$

where

$$\Theta^{\theta' k' M'''}_{\theta k M''}(M, M') = \sum_{\mathbf{k}\mathbf{k}'} \sum_{\sigma\sigma'} \langle \theta \mathbf{k}\sigma | \theta k M''\rangle \times \langle \theta k' M''' | \theta \mathbf{k}'\sigma'\rangle J^{MM'}_{\theta \mathbf{k}\sigma, \theta' \mathbf{k}'_{\sigma}}. \tag{12}$$

The only restriction which is imposed on the symmetry of the exchange interaction is the condition that the interaction (11) have the symmetry of the point group of the impurity site.[31] Generally speaking, the seed basis of crystal-field states $\{|M\rangle\}$ obtained with consideration of only the single-particle crystal potential is not diagonal when the perturbation (11) is taken into account. In low-symmetry systems this perturbation can mix seed states of the crystal-field basis.[32] Therefore, in the general case the relation

$$\Theta^{\theta' k' M'''}_{\theta k M''}(M, M') = \tilde{J}^{MM'}_{\theta k, \theta' k'} \delta_{M''' M} \delta_{M'' M'}, \tag{13}$$

which reduces the exchange Hamiltonian to the standard Coqblin–Schrieffer expression in the partial-wave representation

$$\mathscr{H}_{\rm ex} = \sum_{MM'} f^{\dagger}_{M} f_{M'} \sum_{kk'} \sum_{\theta\theta'} \tilde{J}^{MM'}_{\theta k, \theta' k'} c^{\dagger}_{\theta k M'} c_{\theta' k' M}, \tag{14}$$

is an artefact of the simplifying assumption that the mixing potential has spherical symmetry in the vicinity of the impurity. Nevertheless, even in the simplest approximation, in which the band index θ and the dependence on the wave number k are neglected (i.e., the band system of the conduction electrons is replaced by an effective density of states), the approximate Hamiltonian

$$\mathscr{H}_{\rm ex} = \sum_{MM'} \bar{I}^{MM'} f^{\dagger}_{M} f_{M'} c^{\dagger}_{M'} c_{M}, \tag{15}$$

which faithfully takes into account the principal features of the symmetry of the states of the delocalized electrons, differs significantly from the formal sf exchange Hamiltonian (1). When the relaxation width is calculated, the sf exchange Hamiltonian (1) induces only transitions with a change in the projection of the total angular momentum of the impurity by unity or without any change in its projection. The relative values of the matrix elements specifying the transitions $|M'\rangle \to |M\rangle$ do not depend on the features of the electronic structure and are determined only by the properties of the Pauli matrices and the structure of the wave functions $|M\rangle$ of the localized states. Conversely, all the quantities appearing in the Hamiltonian (11) can be calculated for a specific impurity in a specific crystal, and the parameters of the approximate Hamiltonian (15) are obtained by averaging (11). Thus, in the general case the Hamiltonian (15) has nonzero matrix elements for the transition between any local states $|M\rangle$ and $|M'\rangle$, and the relations between the different matrix elements $\bar{I}^{MM'}$ are determined by the localized states of both the crystal field of the impurity and the band structure of the conduction electrons.

The calculation of the averaged parameters $\bar{I}^{MM'}$ can be performed by the methods described in Refs. 22–25 and is

beyond the scope of the present work. In this paper we wish to analyze which features of the temperature dependence of the relaxation widths of the crystal-field levels can be observed for various relations between the symmetrized exchange constants $\overline{I}^{MM'}$ of the Hamiltonian (15).

3. RELAXATION WIDTHS OF CRYSTAL-FIELD LEVELS AND THEIR INFLUENCE ON THE WIDTHS OF THE PEAKS FOR NEUTRON TRANSITIONS

The relaxation width Γ_{if} associated with the transition $|i\rangle \to |f\rangle$ is determined by the natural widths γ_i and γ_f of the initial $|i\rangle$ and final $|f\rangle$ states. It should be noted that the natural widths are determined not only by the mutual relaxation processes of the initial and final states $|i\rangle \leftrightarrow |f\rangle$, but also by the processes $|i\rangle \leftrightarrow |M\rangle$ ($|f\rangle \leftrightarrow |M\rangle$), which are associated with the interaction of the initial (final) states with all the other crystal-field states $\{|M\rangle\}$. In this case the natural width of the initial (final) state is determined by the set of parameters $\{\overline{I}^{iM}\}$ ($\{\overline{I}^{fM}\}$) of the Hamiltonian (15).

Let us consider the process responsible for the inelastic neutron transition $|i\rangle \to |f\rangle$ from the initial state $|i\rangle$ with the energy E_i to the final state with the energy $E_f = E_i + \Delta_{fi}$. We introduce the Matsubara Green's functions describing the crystal-field states of the impurity center j and the Green's functions of similar nature for Abrikosov pseudofermions,[33]

$$\mathcal{G}_\lambda = -\langle T_\tau f_{j,M}(\tau) f^\dagger_{j,M}(0)\rangle, \tag{16}$$

which have the following forms in the zeroth approximation (i.e., in the absence of relaxation):

$$\mathcal{G}_i^{(0)} = (i\omega - E_i + \mu)^{-1}, \tag{17}$$

$$\mathcal{G}_f^{(0)} = (i\omega - E_i - \Delta_{fi} + \mu)^{-1} \tag{18}$$

(in the notation adopted μ is the chemical potential of the pseudofermions, and in the final formulas it must be assumed that $\mu \to -\infty$).

The retarded Green's functions, which specify the spectral response of the system, can be obtained using the analytic continuation of the Matsubara Green's functions from the upper semiaxis onto the entire complex plane of ω. Passage to the retarded Green's functions in the zeroth-order Green's functions requires the replacement $i\omega \to \omega + i\delta$. The interactions of the crystal-field states with other subsystems of elementary excitations of the crystal lead to renormalization of the crystal-field energy and to the appearance of a frequency-dependent imaginary part in the denominator of the Green's function. The renormalizations of the crystal-field splittings can be included in the definition of the Green's functions (17) and (18) and will not be considered further. Let us next concentrate our attention on the temperature dependence of the relaxation width and take into account that the retarded Green's functions of the crystal-field levels can be written in the pole approximation in the form

$$\mathcal{G}_i^R(\omega) = [\omega - E_i + \mu + i\gamma_i(\omega)]^{-1}, \tag{19}$$

$$\mathcal{G}_f^R(\omega) = [\omega - E_i - \Delta_{fi} + \mu + i\gamma_f(\omega)]^{-1}. \tag{20}$$

The on-site susceptibility, which determines the magnetic neutron response of an impurity center, is expressed[33] in terms of the retarded Green's function

$$\chi_{if}^R(\tau) = -\langle T_\tau f_i^\dagger(\tau) f_f(\tau) f_f^\dagger(0) f_i(0)\rangle,$$

whose analytic continuation onto real frequencies has the following form:

$$\chi_{if}^{(0)R}(\Omega) = -|\Theta_{if}|^2 \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} \tanh\left(\frac{\epsilon}{2T}\right)[\mathrm{Im}\ \mathcal{G}_i^R(\epsilon) \times \mathcal{G}_f^R(\epsilon+\Omega) + \mathrm{Im}\ \mathcal{G}_f^R(\epsilon)\mathcal{G}_i^A(\epsilon-\Omega)] \tag{21}$$

(here Θ_{if} is a matrix element, which depends on the wave functions of the initial and final crystal-field states and determines the intensity of the neutron scattering peak). Representing the resonant part of the susceptibility $\chi_{if}^R(\Omega)$ at $\Omega \approx \Delta_{fi}$ in the form

$$\chi_{if}^R(\Omega) = \frac{\Xi_0}{\Omega - \Delta_{fi} + i\Gamma_{if}}, \tag{22}$$

where Ξ_0 is the residue at the respective pole), we can obtain the dependence of Γ_{if} on the corresponding natural dampings of the pseudofermion Green's functions. In the limit $\gamma_{i,f} \ll \Delta_{fi}$ or $\gamma_{f,i} \ll T$ the relation between the relaxation constant Γ_{if} extracted from the results of magnetic inelastic neutron scattering experiments and the natural damping of the pseudofermion Green's functions acquires a simple form:

$$\Gamma_{if} = \gamma_i(\omega = E_i) + \gamma_f(\omega = E_f). \tag{23}$$

Thus, in the cases which are most interesting for a reliable experimental analysis (where the width of the inelastic transition is smaller than its energy) the problem of determining the temperature dependence of the width Γ_{if} of a transition reduces to a calculation of the natural widths of the initial and final states.

Let us consider the influence of conduction electrons on the natural width of crystal-field states in the Cornut–Coqblin model. For this purpose we use the effective Hamiltonian (15) obtained in the preceding section as the interaction Hamiltonian. The natural widths are calculated by standard Feynman-diagram techniques at finite temperatures. This allows us to partially sum diagram series and to obtain a closed system of self-consistent equations. The departure from perturbation theory is critical in the case of fairly strong relaxation, since the natural width $\gamma_M(\omega = E_M)$ of each crystal-field state $|M\rangle$ depends on the relaxation widths of the entire system of crystal-field levels and must, therefore, be found self-consistently. To illustrate this point, we consider the interaction between the states $|M\rangle$ and $|M'\rangle$ with the energies E_M and $E_{M'} = E_M + \Delta_{M'M}$, respectively. The simplest diagram which leads to relaxation of the crystal-field states is shown in Fig. 1a. The dashed line correspond to the Green's function of the conduction electron

$$G(\mathbf{r},\tau) = -\langle T_\tau \Psi_\xi(\mathbf{r},\tau)\Psi_\xi^\dagger(0,0)\rangle, \quad \xi = M, M' \tag{24}$$

(we neglect the difference between the Green's functions of the conduction electrons for different M). The diagrams corresponding to the vertex corrections can be classified in the following manner. The first are parquet diagrams, which are

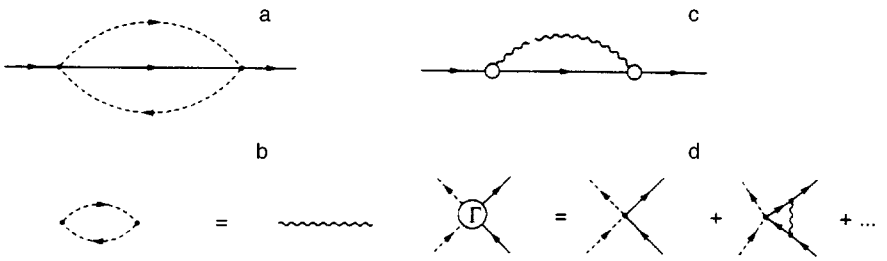

FIG. 1. Feynman diagrams: a—simple diagram describing the shift and damping of a crystal-field level (dashed line—conduction electron Green's function, solid line–crystal-field excitations); b—conduction electron polarization operator, which describes the electromagnetic interaction between crystal-field excitations; c—eigenenergy part of the crystal-field excitations with consideration of the vertex renormalization; d—skeletal diagrams for vertex corrections.

similar to the Abrikosov diagrams considered in the analysis of the Kondo effect in Ref. 33. Consideration of the contribution from the first nonvanishing term leads to the appearance of an interaction in the channel $\Gamma_{M=M'} \sim (I^2_{MM'}/W)\ln(W/\Delta_{MM'})$ and to the corrections $\delta\Gamma^{(3,p)}_{MM'} \sim (I^3_{MM'}/W^2)\ln^2(W/\Delta_{MM'})$ (W is the width of the conduction electron band). The second are nonparquet diagrams.[34] Consideration of the contribution from the first correction (Fig. 1d) leads to the additional contribution $\delta\Gamma^{(3,np)}_{MM'} \sim (I^3_{MM'}/W^2)\ln(W/\Delta_{MM'})$. We shall henceforth assume $I_{MM'}/W \ll 1$ and $(I_{MM'}/W)\ln(W/\Delta_{MM'}) \ll 1$ and neglect the vertex corrections in the perturbative approach. Under these circumstances

$$\sigma_M(i\omega_n) = [\overline{I}^{MM'}]^2 T^2 \sum_{\epsilon_1\epsilon_2} \frac{1}{\mathscr{N}^2} \sum_{\mathbf{p}_1,\mathbf{p}_2} G(\mathbf{p}_1,\epsilon_1)G(\mathbf{p}_2,\epsilon_2) \times \mathscr{G}_{M'(\neq M)}(\epsilon_1+\epsilon_2-\omega) \tag{25}$$

($\mathscr{N}$ is the total number of conduction electrons). Performing the analytic continuation of the expression (25) into the upper half-plane of the complex variable ω according to the usual rules,[35–37] we obtain the following expressions for the eigenenergy parts at real frequencies (the analogous equations for the sf exchange Hamiltonian were obtained by Maleev in a treatment of the relaxation of the crystal field in cubic metals[19]):

$$\gamma_M(\omega) = -\,\mathrm{Im}\ \sigma^R_M(\omega) = \frac{1}{\pi}[\overline{I}^{MM'}]^2 \int_{-\infty}^{\infty} dx \frac{1}{\mathscr{N}} \sum_{\mathbf{p}} [N(x)+n(x+\omega)]\,\mathrm{Im}\ \mathscr{G}^R_{M'(\neq M)}(x+\omega)\,\mathrm{Im}\ \Pi^R_{MM'}(\mathbf{p},x). \tag{26}$$

Here $N(x)=(e^{x/T}-1)^{-1}$, $n(x)=(e^{x/T}+1)^{-1}$, $\Pi(\mathbf{p},x)$ is the polarization operator of the conduction electrons (Fig. 1b), whose imaginary part describes the two-particle density of states:

$$\mathrm{Im}\,\frac{1}{\mathscr{N}}\sum_{\mathbf{p}} \Pi^R(\mathbf{p},x) = -\frac{\pi}{2}N_0^2 x, \tag{27}$$

where N_0 is the single-particle density of states of the conduction electrons at the Fermi level, in terms of which the dimensionless coupling constants $g_{MM'}$ are expressed:

$$g^2_{MM'} = \frac{1}{2}[I^{MM'}N_0]^2. \tag{28}$$

In the integrals (26) we perform the replacement $\omega+\mu=\tilde{\omega}$ corresponding to the displacement of the energy reference point. Allowing μ to tend to $-\infty$, we neglect the Fermi function on the right-hand side. This replacement has a simple physical meaning: the singularities of the functions $\mathscr{G}$ are determined by a far larger energy scale, and, therefore, the terms corresponding to consideration of the poles of the pseudofermion functions should be omitted.[38]

According to (23), the natural damping γ of the crystal-field states at the frequencies corresponding to the energies of the crystal-field levels must be calculated to determine the width of a neutron transition. Thus, in the case of the interaction of $|M\rangle$ and $|M'\rangle$, the quantities $\gamma_M(\omega=E_M)$ and $\gamma_{M'}(\omega=E_{M'})$ must be calculated. Determining the damping at the poles of the corresponding Green's functions, we obtain the system of coupled equations

$$\begin{cases} \gamma_M(\omega\to E_M) = \pi g^2_{MM'} \int_{-\infty}^{\infty} dx x N(x) P \times (x-\Delta_{M'M},\gamma_{M'}), \\ \gamma_{M'}(\omega\to E_M+\Delta_{M'M}) = \pi g^2_{MM'} \int_{-\infty}^{\infty} dx x N(x) \times P(x+\Delta_{M'M},\gamma_M), \end{cases} \tag{29}$$

where $P(x,\gamma)$ is the spectral function normalized to unity:

$$P(x,\gamma) = \frac{1}{\pi}\frac{\gamma}{x^2+\gamma^2}. \tag{30}$$

The expressions obtained are easily generalized to the case of an arbitrary set of constants in the Hamiltonian (15) ($\{\overline{I}^{MM'}\}$; $M,M'=1,\ldots,2J_{\mathrm{imp}}+1$) and an arbitrary system of crystal-field states with the energies E_M. Proceeding precisely as in the derivation of (25)–(29), we obtain the expressions for the frequency-dependent damping rates

$$\gamma^R_M(\omega) = -\,\mathrm{Im}\ \sigma^R_M(\omega) = \frac{1}{\pi}\sum_{M'=1}^{2J_{\mathrm{imp}}+1} \overline{I}^2_{MM'} \int_{-\infty}^{\infty} dx \frac{1}{\mathscr{N}} \sum_{\mathbf{p}} [N(x)+n(x+\omega)]\,\mathrm{Im}\ \mathscr{G}^R_{M'}(x+\omega)\,\mathrm{Im}\ \Pi^R_{MM'}(\mathbf{p},x),$$
$$M=1,\ldots,2J_{\mathrm{imp}}+1. \tag{31}$$

Neglecting the Fermi function on the right-hand side of (31), we obtain the system of self-consistent equations

$$\gamma_M(\omega) = -\sum_{M'=1}^{2J_{\rm imp}+1} g^2_{MM'} \int_{-\infty}^{\infty} dx x N(x) \mathrm{Im}\ \mathcal{G}^R_{M'}(x+\omega),$$

$$M = 1,\ldots,2J_{\rm imp}+1. \tag{32}$$

Here the dimensionless coupling constants are expressed in terms of the parameters of the Hamiltonian (15):

$$g^2_{MM'} = g^2_{M'M} = \frac{1}{2}[\overline{I}^{MM'} N_0^{MM'}]^2, \tag{33}$$

where $N_0^{MM'}$ is the partial density of states of the conduction electrons corresponding to the $M \to M'$ transition.

The system of equations for finding the natural relaxation constants at the frequencies which determine the widths of the neutron transitions [see (23)] can be represented in the explicit form[2)]

$$\gamma_M(\omega \to E_M) = \int_{-\infty}^{\infty} dx x N(x) \times \sum_{M'=1}^{2J_{\rm imp}+1} g^2_{MM'} P(x - \Delta_{M'M}, \gamma_{M'}),$$

$$M = 1,\ldots,2J_{\rm imp}+1, \tag{34}$$

where

$$\Delta_{M'M} = E_{M'} - E_M.$$

4. CLASSIFICATION OF THE TEMPERATURE DEPENDENCE OF RELAXATION WIDTHS

The temperature dependence of the natural relaxation widths (and the widths of the neutron transitions determined by them) depends on the relationship between the different constants in the Hamiltonian (15) and on the energies of the crystal-field states. In this section we shall classify the types of temperature dependence for cases in which solutions of the self-consistent system of equations (34) can be obtained explicitly.

The simplest condition under which the system of equations (34) is decoupled is that the relaxation widths be small ($\gamma_M \to 0$). In this case, instead of the system of equations (32), we obtain the following expressions for the non-self-consistent widths of the levels $\gamma_M^{(0)}$:

$$\gamma_M^{(0)} = \pi \sum_{M'=1}^{L} \int_{-\infty}^{\infty} g^2_{MM'} x N(x)\, \delta(x - \Delta_{MM'}). \tag{35}$$

The calculation of (35) permits separation of the contributions to the temperature dependence of the natural width $\gamma_M(T)$ into three types:

$$\gamma_M(T) = \gamma_M^{(\rm eq)}(T) + \gamma_M^{\uparrow}(T) + \gamma_M^{\downarrow}(T). \tag{36}$$

The first type is associated with the relaxation caused by the interaction of the crystal-field state $|M\rangle$ with the levels $\{|M'\rangle\}$, whose energies $E_{M'}$ equal E_M:

$$\gamma_M^{(\rm eq)}(T) = \pi T \sum_{M'}^{E_{M'}=E_M} g^2_{MM'}. \tag{37}$$

The contributions to the natural relaxation width of the level $|M\rangle$ from higher-lying ($E_{M'} > E_M$) and lower-lying ($E_{M'} < E_M$) levels are given by the expressions

$$\gamma_M^{\uparrow}(T) = \pi \sum_{M'}^{E_{M'}>E_M} g^2_{MM'} \Delta_{M'M} N(\Delta_{M'M}) \tag{38}$$

and

$$\gamma_M^{\downarrow}(T) = \pi \sum_{M'}^{E_{M'}<E_M} g^2_{MM'} \Delta_{MM'} [N(\Delta_{MM'}) + 1], \tag{39}$$

respectively. In the limit of high temperatures, $T \gg \max(E_M)$, in accordance with the results in Refs. 17–21, the temperatures dependences of all three contributions to the relaxation are indistinguishable. All three contributions obey a Korringa law, and the expression for the natural relaxation width takes the form

$$\gamma_M(T) = \pi T \sum_{M'=1}^{2J_{\rm imp}+1} g^2_{MM'}. \tag{40}$$

At low temperatures the contributions of the higher-lying ($E_{M'} > E_M$) and lower-lying ($E_{M'} < E_M$) crystal-field levels differ significantly. In the limit $T \to 0$, $N(\Delta)$ is exponentially small, $N(\Delta) \to \exp(-\Delta/T)$, and (38)–(39) take the form

$$\gamma_M^{\uparrow}(T) = \pi \sum_{M'}^{E_{M'}>E_M} g^2_{MM'} \Delta_{M'M} \exp\left(-\frac{\Delta_{M'M}}{T}\right), \tag{41}$$

$$\gamma_M^{\downarrow}(T) = \pi \sum_{M'}^{E_{M'}<E_M} g^2_{MM'} \Delta_{MM'}. \tag{42}$$

Thus, as $T \to 0$, the contributions to the natural relaxation width from the higher-lying levels tend exponentially to zero, and the contributions from the lower-lying levels do not depend on the temperature.

Since the shape of the line for the neutron transition $|i\rangle \to |f\rangle$ is measured directly in an experiment, it would be interesting to analyze the temperature dependence of the width of the transition $\Gamma_{i\to f}(T) = \gamma_i(T) + \gamma_f(T)$ (see Fig. 2) for different relationships between the constants of the Hamiltonian (1). The diagonal interactions I^{ii} (I^{ff}) lead to a contribution $\sim \pi g_{ii}^2 T$ ($\sim \pi g_{ff}^2 T$), which is proportional to the temperature. In the case of the relaxation of only the initial (final) state as a result of interactions with the upper ($\uparrow$) levels, we have $[\Gamma^{\uparrow}_{i\to f}(T)]^{i(f)} \sim N(\Delta_{\uparrow i(f)})$, which leads to exponentially small damping, $\sim \exp(-\Delta_{\uparrow i(f)}/T)$, at low temperatures. When only the initial (final) state relaxes as a result of interactions with lower ($\downarrow$) levels, we have $[\Gamma^{\downarrow}_{i\to f}(T)]^{i(f)} \sim \Delta_{i(f)\downarrow}[N(\Delta_{i(f)\downarrow}) + 1]$, which can be described by a constant $\sim \Delta_{i(f)\downarrow}$ at low temperatures. The fourth special case is the one in which relaxation is mediated by the interaction between the initial and final states I^{if}. In this case $\Gamma^{if}_{i\to f}(T) \sim \Delta_{fi}[2N(\Delta_{fi}) + 1] = \Delta_{fi} \coth(\Delta_{fi}/2T)$.

If the special cases just described are realized in the system being studied, they are easily distinguished from one another even by qualitative visual inspection. The situation in which relaxation of the initial and final levels occurs only because of the influence of the higher-lying levels is easily distinguished (the dotted line in Fig. 3). In this case

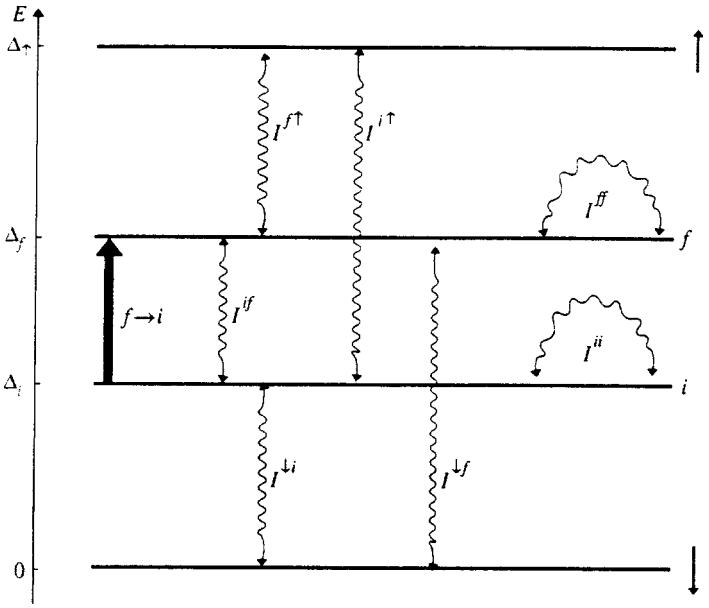

FIG. 2. General case of the classification of sources for the relaxation of the levels of the initial (i) and final (f) states of a transition (thick vertical arrow) due to interactions with lower-lying levels ($I^{\downarrow i}$ and $I^{\downarrow f}$) and higher-lying levels ($I^{\uparrow i}$ and $I^{\uparrow f}$) and due to mutual coupling of the initial and final states (I^{if}, wavy lines).

$\Gamma^{\uparrow}_{i\to f}(T\to 0)\to 0$. The mutual relaxation processes (the solid line in Fig. 3) are also visually distinguishable from the variants in which the broadening is a consequence of the interactions of the initial or final state with lower-lying levels (the dashed line in Fig. 3). The sharpness of the temperature dependence can serve as a criterion in these cases. In the former variant (see Fig. 3) we have $\Gamma^{if}_{i\to f}(T=2\Delta_{MM'})/\Gamma^{if}_{i\to f}(T=0)\approx 4$, and in the latter variant we have $\Gamma^{\downarrow}_{i\to f}(T=2\Delta_{MM'})/\Gamma^{\downarrow}_{i\to f}(T=0)\approx 2.5$.

5. CONSEQUENCES OF THE SELF-CONSISTENT PROCEDURE

Beside the obvious quantitative influence of the self-consistent procedure manifested as renormalization of the numerical values of the natural relaxation constants, there is a qualitative difference, which is expressed by the nonzero value of the relaxation width of the ground state $|G\rangle$ at zero temperature.

In the non-self-consistent procedure [see (41)] the width $\gamma^{(0)}_G(T\to 0)=\pi g^2_{GE}\Delta_{EG}\exp(-\Delta_{EG}/T)\to 0$ (E is the higher-lying level with the smallest value of Δ_{EG}). The solution of the system of self-consistent equations (29) (for $M=G$, $M'=E$) gives a nonzero width: $\gamma^{(sc)}_G(T=0)\neq 0$. Under the conditions $T\ll\gamma_{EG}$ and $T\ll\Delta_{EG}$ an explicit expression can be obtained for the width $\gamma^{(sc)}_G(T=0)$ of the level. Since at low temperatures the non-self-consistent natural width of the crystal-field excited level is $\gamma^{(0)}_E(T\to 0)=\pi g^2_{GE}\Delta_{EG}$, the weak corrections caused by the influence of the lower level can be neglected. Then the self-consistent width of the lower level is proportional to the square of the coupling constant:

$$\gamma^{(sc)}_G\approx\pi g^2_{GE}\gamma^{(0)}_E\ \ln\left(\frac{W}{\Delta_{GE}}\right) \tag{43}$$

(in the calculation we cut off the integral (29) at the width of the conduction electron band W). Substituting the expression for $\gamma^{(0)}_E$ into (43), we obtain[3)]

$$\gamma^{(sc)}_G\approx\pi^2 g^4_{GE}\Delta_{EG}\ \ln\left(\frac{W}{\Delta_{GE}}\right). \tag{44}$$

Since the corrections associated with the influence of the width of the lower level on the upper level contain an additional small factor $\sim g^2_{GE}$, (43) is the explicit solution of the system of self-consistent equations to within terms $\sim g^4_{GE}$ inclusively.

This result, which is unexpected from the standpoint of perturbation theory, can have a physical interpretation in the self-consistent theory. It should, first of all, be taken into account that $|G\rangle$ is the ground state of the system only when the interactions are disregarded. When the interaction with the delocalized conduction electrons is included, the nomenclature of the localized states is no longer the true quantum-mechanical basis, and $|G\rangle$ is not the true ground state.

A specific mechanism, which causes damping of the crystal-field state $|G\rangle$ at zero temperature, can be pointed out. The physical cause of the damping of $|G\rangle$ is the nonzero broadening of the excited state $|E\rangle$ (which also occurs in perturbation theory). Figure 4 presents the spectral functions of $|E\rangle$ [$P(x-1,\gamma^{(0)}_E=0.4)$] and $|G\rangle$ [$P(x,\gamma^{(0)}_G=0)=\delta(x)$] in the perturbative approximation. The width of $|G\rangle$ in the

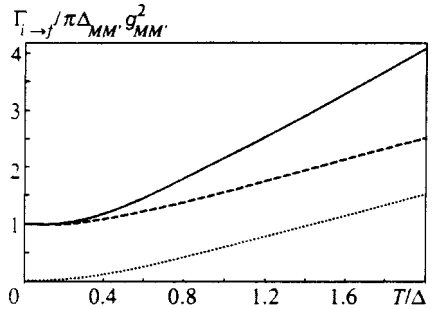

FIG. 3. Reduced temperature dependences of the total inelastic scattering width Γ_{if} due to relaxation processes with a higher-lying level (dotted line) and a lower-lying level (dashed line) and mutual processes (I^{if}, solid line).

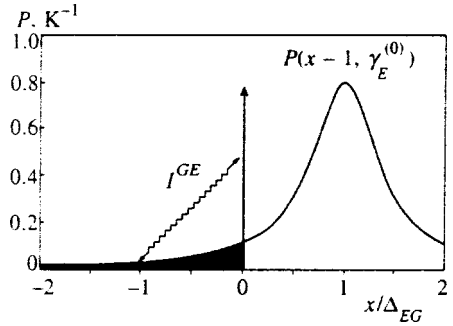

FIG. 4. Illustration of the origin of the finite width of the crystal-field ground state $|G\rangle$ (vertical arrow at $x/\Delta_{EG}=0$) due to relaxation processes (wavy lines with arrows) which couple the $|G\rangle$ level to the low-energy tail (darkened region) of the spectral function $P(x-1,\gamma_E^{(0)}\neq 0)$ of the upper $|E\rangle$ state.

self-consistent approach is nonzero ($\gamma_G^{(\mathrm{sc})}\neq 0$) because of the allowed transitions I^{GE} induced by the width of the upper level $|E\rangle$ to the low-energy tail (the darkened area in Fig. 4) of the Lorentzian contour of the upper level.

The result obtained, which attests to the nonzero contribution to the natural relaxation width of the ground-state level, calls for caution in approaching methods for estimating the characteristic temperature T^* in Kondo systems from the full width at half maximum (FWHM) of the quasielastic neutron scattering peak at zero temperature. According to the generally accepted approach,[39,40] the characteristic temperature is determined from the relation $T^*=\Gamma_{\mathrm{qe}}^{\mathrm{exp}}(T=0)/2$, where $\Gamma_{\mathrm{qe}}^{\mathrm{exp}}(T=0)$ is the experimentally observed quasielastic scattering width at zero temperature. In this procedure it is assumed that the width of the peak $\Gamma_{\mathrm{qe}}^{\mathrm{exp}}(T=0)$ is determined only by the anomalous width $\Gamma_K(T=0)$, which is associated with Kondo scattering processes on the lowest crystal-field state. However, the presence of the nonzero contribution $\gamma_G^{(\mathrm{sc})}(T=0)$ from the normal relaxation processes calls for additional refinement in the case of strong relaxation broadening in systems with soft crystal fields. In this situation, since the experimental width $\Gamma_{\mathrm{qe}}^{\mathrm{exp}}(T=0)$ is determined not only by the anomalous width $\Gamma_K(T=0)$, but also by the relaxation contribution $\gamma_G^{(\mathrm{sc})}(T=0)$, we have

$$\Gamma_{\mathrm{qe}}^{\mathrm{exp}}(T=0)=\Gamma_K(T=0)+2\gamma_G^{(\mathrm{sc})}(T=0), \tag{45}$$

and the standard relation should be rewritten in the form

$$T^*=\frac{\Gamma_{\mathrm{qe}}^{\mathrm{exp}}(T=0)-2\gamma_G^{(\mathrm{sc})}(T=0)}{2}. \tag{46}$$

Thus, when there is strong relaxation in systems with soft crystal fields, the determination of the characteristic temperature is complicated by the nonzero relaxation contribution at $T=0$. Nevertheless, the use of (46) and (44) provides an estimate in this case too. To analyze the contribution from the normal relaxation processes of a specific compound, we must determine the parameters (the crystal-field splitting $\Delta_{MM'}$ and the dimensionless relaxation constants $g_{MM'}$) which describe the relaxation in the particular material. The set of techniques discussed in the next section can be useful in solving this problem.

6. PARAMAGNETIC LABELING

Studying the electronic structure by measuring the relaxation of a paramagnetic label requires the observance of several conditions, which must be satisfied by the compound being studied $A_l\{B\}$ and the paramagnetic ion P. In this section we describe the most desirable general conditions, under which performing and interpreting paramagnetic labeling experiments are simplest, and we present some examples of compounds which satisfy these necessary conditions.

One necessary condition which must be satisfied by the compound $A_l\{B\}$ being studied is the existence of a reference compound $P_l\{B\}$ containing the paramagnetic label P. The reference compound must be a structural analog of the compound being studied. In the first stage the properties of the reference compound $P_l\{B\}$ must be investigated. The purpose of studying the reference compound is to obtain information on the crystal-field energies and wave functions of the paramagnetic label. This information can be obtained by analyzing experimental data from measurements of magnetic neutron scattering (or Raman scattering) and the thermodynamic properties (the magnetic susceptibility and specific heat). For neutron scattering experiments, which require a fairly large quantity of the material, it should be noted that a single-crystal sample is not required. This greatly facilitates implementation of the method, since magnetic susceptibility data suitable for reconstructing the crystal-field wave functions can be obtained from measurements on tiny single-crystal samples.

In the second stage, for which a polycrystalline sample suffices, inelastic neutron scattering experiments are performed on the compound $(A_{1-x}P_x)_l\{B\}$. The theoretical analysis requires information on the crystal-field states of the paramagnetic label P in $(A_{1-x}P_x)_l\{B\}$. The experimental neutron scattering data provide information on the energies of the crystal-field levels of the paramagnetic label in the compound being studied. Since it is impossible to study the crystal-field states of an ion of P in $(A_{1-x}P_x)_l\{B\}$ by thermodynamic methods, additional information on the crystal-field wave functions is needed. This information can be obtained by studying the trends in the variation of the crystal-field parameters of a family of compounds $R_l\{B\}$ (where R=A,P,...). There are presently several families of compounds for which such investigations have already been performed: RAl_3,[9–11] RNi_5,[12–14], and RNi[14–16] (R is a rare-earth ion). In these families the main contribution to the formation of the crystal field is made by the ions in the local environment and the conduction electrons. Therefore, the structures of the crystal-field wave functions of the paramagnetic label P in the reference compound and in the compound being studied are practically identical. Thus, the systems which are suitable for the proposed procedure are compounds in which the nearest neighbors of each ion of A that is replaced by the paramagnetic label are ions from an unsubstituted formula unit of {B}. The best systems for application of the method are materials in which the crystal-field parameters are determined predominantly by the nearest neighbors.

The next necessary condition is a small concentration of the ions of the paramagnetic label, $x\ll 1$, in $(A_{1-x}P_x)_l\{B\}$.

This condition permits consideration of the crystal fields and their relaxation as purely single-ion phenomena. A small value of x is also necessary to be sure that doping with the paramagnetic label did not lead to significant alteration of the electronic structure of the compound being studied.

An important desirable restriction imposed on the paramagnetic label is nondegeneracy of the crystal-field ground state of the ion. Fulfillment of this condition significantly simplifies the interpretation of the experimental data, since the natural width of the crystal-field ground state is specified by a simple exchange Hamiltonian of the form (15). A degenerate ground state $|G\rangle$ can lead to a Kondo effect, which results in the appearance of a specific temperature dependence of the quasielastic neutron scattering width: $\Gamma_{qe}(T) = 2\gamma_G(T) = a + b\sqrt{T}$.[40] The presence of a specific temperature dependence of the natural width of the crystal-field ground state greatly complicates the interpretation, since the width $\Gamma_{GM}(T) = \gamma_G(T) + \gamma_M(T)$ of any inelastic transition from the ground state contains this poorly studied component. This circumstance greatly complicates the analysis, since the theoretical treatment is simplest, if the temperature dependence of the natural widths $\gamma_M(T)$ of the levels is determined during the experiment. This dependence can be extracted from the solution of the system of equations

$$\Gamma_{MM'}(T) \approx \gamma_M(T) + \gamma_{M'}(T), \tag{47}$$

whose features must be analyzed separately for each specific case.

It is also noteworthy that one of the significant advantages of the method can be the possibility of regulating the selection rules by choosing different paramagnetic labels. It can easily be seen that different symmetries for the crystal-field states of the paramagnetic label will lead to different types of temperature dependence of the relaxation process. Therefore, significant information can be obtained by analyzing situations with different ions serving as the paramagnetic label in relaxation spectroscopy.

7. RELAXATION OF THE PARAMAGNETIC LABEL Pr^{3+} IN $Pr_{0.03}Ce_{0.97}Al_3$ AND $Pr_{0.03}La_{0.97}Al_3$

An example of a favorable combination of properties for the paramagnetic label and the compound being studied is the Pr^{3+} ion in the hexagonal compounds RAl_3 (R is a lanthanide). The crystal-field ground state (level 1 in Fig. 5) in pure $PrAl_3$ is the singlet $|\Gamma_1\rangle = |0\rangle$, and the only allowed transition at $T \to 0$ is the $|\Gamma_1\rangle \to |\Gamma_6\rangle$ transition to the $|\Gamma_6\rangle = |\pm 1\rangle$ state (level 2 in Fig. 5). The crystal fields of the praseodymium ion in $PrAl_3$ were studied in detail in Ref. 9 (see Fig. 5, in which the crystal-field levels are numbered from 1 to 6 in order of increasing energy). The singlet character of the ground state rules out both the Korringa relaxation channel $\sim |I^{11}|^2$ and the relaxation channel associated with Kondo processes.

The relaxation of the paramagnetic label Pr^{3+} was studied in $Pr_{0.03}Ce_{0.97}Al_3$ and $Pr_{0.03}La_{0.97}Al_3$. The crystal-field splitting energy of the praseodymium ion Δ_{21} in both $Pr_{0.03}Ce_{0.97}Al_3$ and $Pr_{0.03}La_{0.97}Al_3$ differs only slightly from the crystal-field energy $\Delta_{21}^R \approx 4.5$ meV in the reference compound $PrAl_3$. More specifically, in $Pr_{0.03}Ce_{0.97}Al_3$ $\Delta_{21} \approx 4.2$ meV, and in $Pr_{0.03}La_{0.97}Al_3$ $\Delta_{21} \approx 3.5$ meV. Therefore, in the simplest approximation the wave functions of the crystal-field states of the paramagnetic label in the compounds studied can be assumed to be only slightly altered from those in the reference. To describe the relaxation of the paramagnetic label in $CeAl_3$, we selected a system of levels which coincides with the crystal-field system in pure $PrAl_3$, and to analyze the relaxation in $LaAl_3$, we chose a system in which all the splitting energies are reduced by a factor of $3.2/4.5 \approx 0.711$ (see Fig. 5). The thoroughly studied laws governing the variation of the crystal field of the paramagnetic label in $Pr(La)Al_3$ (Refs. 9 and 10) can be used for a more exact calculation.

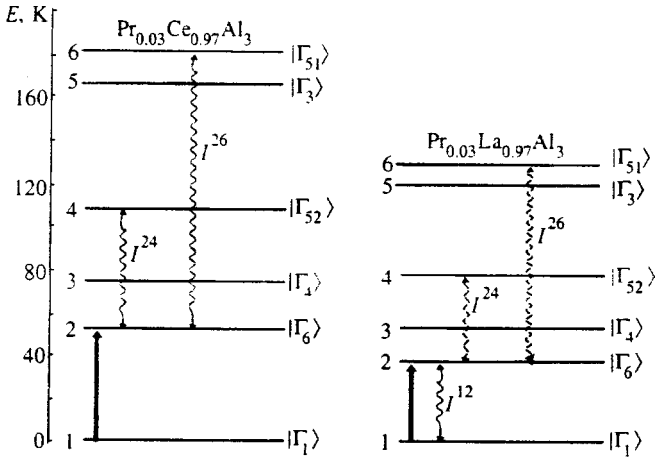

FIG. 5. Level scheme of the paramagnetic label Pr^{3+} in $CeAl_3$ (on the left) and $LaAl_3$ (on the right). Wave functions of the levels:[9] $|\Gamma_1\rangle = |0\rangle$; $|\Gamma_6\rangle = |\pm 1\rangle$; $|\Gamma_4\rangle = 2^{-1/2}|-3\rangle - 2^{-1/2}|-3\rangle$; $|\Gamma_{52}\rangle = a|\pm 4\rangle - \sqrt{1-a^2}|\mp 2\rangle$; $|\Gamma_3\rangle = 2^{-1/2}|3\rangle + 2^{-1/2}|-3\rangle$; $|\Gamma_{51}\rangle = \sqrt{1-a^2}|\pm 4\rangle + a|\mp 2\rangle$. The neutron transition studied in Ref. 9 is denoted by a vertical arrow. The postulated relaxation channels affecting the initial and final states are denoted by wavy lines with arrows.

In the experiments in Ref. 11 measurements of the temperature dependence of only the transition width $\Gamma_{12}(T)$ were performed (the FWHM of the Lorentzian, which corresponds to 2Γ in our notation, was measured in Ref. 11), while the natural relaxation widths $\gamma_1(T)$ and $\gamma_2(T)$ were not distinguished. Nevertheless, even in this case definite conclusions regarding the difference between the relaxation behavior of the paramagnetic label in $CeAl_3$ and $LaAl_3$ can be drawn.

Since level 1 of the paramagnetic label in $CeAl_3$ corresponds to the ground state, the relaxation of level 1 in interactions with lower-lying levels is impossible. Moreover, a visual comparison of the experimental data (Fig. 6) with the calculated dependences shown in Figs. 2 and 3 allows us to state that the mutual relaxation processes I^{12} are also absent. This conclusion can be drawn on the basis of a comparison of the widths at low and high temperatures: there is no temperature-dependent contribution at $T < 20$ K. Therefore, the only possible sources of natural relaxation broadening of levels 1 and 2 are the interactions of levels 1 and 2 with higher-lying levels 3, 4, 5, and 6.

Although the only quantum numbers in whose nomen-

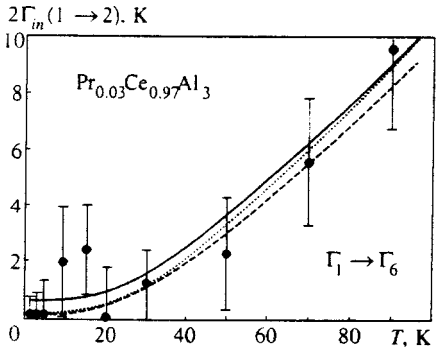

FIG. 6. Temperature dependence of the FWHM ($2\Gamma_{in}$) of the inelastic $|\Gamma_1\rangle\rightarrow|\Gamma_6\rangle$ transition in $CeAl_3$ for the paramagnetic label Pr using the level scheme in Fig. 5. Dotted line—best fitting in the non-self-consistent approximation for $g=g^{(0)}=0.115$. Solid line—best fitting in the self-consistent approximation for $g=g^{(sc)}=0.111$. Dashed line—non-self-consistent width for $g=g^{(sc)}=0.111$.

clature correct arguments can be advanced are the indices of the irreducible representations $M=\Gamma_1,\Gamma_3,\Gamma_4,\Gamma_6,\Gamma_{51}$, and Γ_{52}, we shall demonstrate that the nomenclature of the projections m of the spherical representation [see (5)] is also useful for a qualitative analysis of relaxation.

Using the known wave functions of the crystal-field states (see the caption to Fig. 5), we can rule out the transitions with $\delta m=\pm 1$ because of the lack of the $|\Gamma_1\rangle\leftrightarrow|\Gamma_6\rangle$ mutual relaxation channel. The transitions with $\delta m=\pm 2$ should lead to interactions of $|\Gamma_1\rangle$ with $|\Gamma_{51}\rangle$ and $|\Gamma_{52}\rangle$ and of $|\Gamma_6\rangle$ with $|\Gamma_4\rangle$ and $|\Gamma_3\rangle$. However, if the occurrence of transitions with $\delta m=\pm 2$ is assumed, the relaxation of $|\Gamma_6\rangle$ according to a Korringa law ($\propto T$) should be observed. Then (if it is assumed within a qualitative analysis that the interaction constants are identical for all the transitions with $\delta m=\pm 2$) the occurrence of relaxation according to a Korringa law does not correspond to the weak dependence of Γ_{12} on the temperature in the range $0<T<20$ K (see the experimental points in Fig. 6). Therefore, the occurrence of transitions with $\delta m=\pm 2$ should also be ruled out.

The next possible change in the spherical projection, $\delta m=\pm 3$, leads to interactions of $|\Gamma_1\rangle$ with $|\Gamma_4\rangle$ and $|\Gamma_3\rangle$ and of $|\Gamma_6\rangle$ with $|\Gamma_{52}\rangle$ and $|\Gamma_{51}\rangle$. These interactions do not lead to a contribution that is proportional to T to the widths of the levels of the initial state (γ_1) and the final state (γ_2) and do not contradict the weak temperature dependence of Γ_{12} at $T<20$ K. The energy splittings Δ_{42} and Δ_{62} corresponding to the interactions of the $|\Gamma_6\rangle$ level are smaller than the corresponding splittings Δ_{41} and Δ_{51} for the $|\Gamma_1\rangle$ level. Therefore, the contribution $\propto\exp(-\Delta/T)$ to the transition width Γ_{12} at $T<100$ K (under the assumption of approximately equal values of g for all $\delta m=\pm 3$) from the natural width $\gamma_2(T)$ of the $|\Gamma_6\rangle$ level can be considered the main contribution.

In the quantitative calculations presented below we took into account only the interaction constants I^{24} and I^{26}. In such an approximation the natural width of the crystal-field ground state is $\gamma_1(T)=0$, and, therefore, $\Gamma_{12}=\gamma_2(T)$. For simplicity, the values of I^{24} and I^{26} were set equal to one another (see Fig. 5). The conduction electron band was approximated by a constant density of states with a width of 2 eV. The best fit for the experimental data in the non-self-consistent approximation is achieved with the value of the dimensionless coupling constants $g^{(0)}=g^{(0)}_{|\Gamma_6\rangle|\Gamma_{52}\rangle}\equiv g^{(0)}_{|\Gamma_6\rangle|\Gamma_{51}\rangle}=0.115$. The self-consistent procedure gives the best results when $g^{(sc)}=g^{(sc)}_{|\Gamma_6\rangle|\Gamma_{52}\rangle}\equiv g^{(sc)}_{|\Gamma_6\rangle|\Gamma_{51}\rangle}=0.111$. It is noteworthy that the self-consistent value of the natural width $\gamma_{\Gamma_6}(T=0)$ obtained in the numerical calculation coincides to within a few percent with the results of the analytical formula (44). Figure 6 presents a comparison of the theoretical temperature dependences of the width of the $1\rightarrow 2$ inelastic neutron transition with experimental data. The theoretical data are presented in different approximations, viz., the self-consistent and non-self-consistent approximations. To illustrate the influence of the self-consistent approximation, the figure shows the temperature dependences of the self-consistent and non-self-consistent widths (the solid and dashed lines, respectively) calculated for the same dimensionless constant $g^{(sc)}=0.111$. It is seen from Fig. 6 that the self-consistent width is greater than the non-self-consistent width for the same value of the interaction constant. The effect of the self-consistent approximation, $\Gamma^{(sc)}_{12}/\Gamma^{(0)}_{12}\gg 1$, is most easily observed when the non-self-consistent width is small. Unfortunately, the instrumental errors of the neutron scattering method make it difficult to reliably isolate the effects of the self-consistent approximation. Therefore, the performance of Raman scattering experiments, whose experimental errors are considerably smaller, can provide important additional information.

A qualitative analysis of the experimental temperature dependence of the relaxation in $LaAl_3$ like the analysis performed above for $CeAl_3$ shows that the main relaxation channel corresponds to a change in the spherical projection $\delta m=\pm 1$. We note that this channel does not lead to Korringa relaxation for the initial state 1 or the final state 2. In this approximation relaxation of the initial state is possible only in the mutual I^{12} processes, and relaxation of the final state is possible both in the mutual I^{12} processes and in the I^{24} and I^{26} interactions with higher-lying levels (see the right-hand part of Fig. 5).

Since the non-self-consistent width is greater at all temperatures, it is difficult to observe the effects of the self-consistent approximation. Therefore, we calculated the transition width only in the non-self-consistent approximation (Fig. 7). It is noteworthy that the results of the fitting with consideration of only I^{12} are in better agreement with the experimental data than are the results of the calculations with consideration of equal values for all the interaction constants $I^{12}=I^{24}/\sqrt{1-a^2}=I^{26}/a$. These results are reminiscent of the qualitative character of the estimates based on the nomenclature of the spherical projections of the angular momentum m. A more rigorous approach would take into account, for example, that the contributions of the $m=0\leftrightarrow m=\pm 1$ and $m=\pm 1\leftrightarrow m=\pm 2$ transitions, which are both associated with a change in the spherical projection $\delta m=\pm 1$, can be significantly different in a lattice of fairly low symmetry.

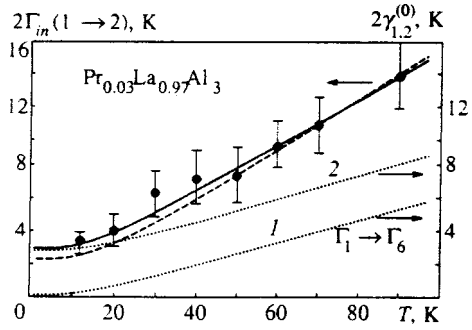

FIG. 7. Temperature dependence of the FWHM ($2\Gamma_{in}$) of the inelastic $|\Gamma_1\rangle \rightarrow |\Gamma_6\rangle$ transition in $LaAl_3$ for the paramagnetic label Pr in the non-self-consistent approximation using the level scheme in Fig. 5. Dashed line—model with $g_{12}=g_{24}/\sqrt{1-a^2}=g_{26}/a=0.094$; solid line—model with $g_{24}=g_{26}=0$, $g_{12}=0.108$. The dotted line *1* (*2*) corresponds to the natural relaxation width $\gamma_{\Gamma_1}(T)$ ($\gamma_{\Gamma_6}(T)$) of the levels.

8. CONCLUSIONS

The proposed self-consistent microscopic theory of the spectroscopy of the crystal-field levels of an impurity ion in a normal metal has several consequences, which can have a significant influence on our understanding of the relaxation processes in highly correlated systems.

Most importantly, the proposed approach, which is based on the Coqblin–Schrieffer–Cooper approach, rather than the formal sf exchange model, permits a first-principles microscopic calculation of the temperature dependence of the relaxation width of crystal-field states. A comparison of such calculations with experimental data, as well as the use of the conventional methods of infrared, Raman, and photoemission spectroscopy, makes it possible to test the faithfulness of the band calculations. Like the traditional methods, the proposed method permits the performance of a quick qualitative visual analysis of the experimental results obtained. An additional significant feature of the proposed spectroscopic method is the possibility of regulating the selection rules by choosing an appropriate paramagnetic label, which is impossible within the traditional methods.

The new soft spectroscopic method considered in this paper permits the investigation of the role of strong electron correlations in shaping the relaxation processes of crystal-field states. Since strong correlations can significantly alter the simple form of the effective Hamiltonian (15), relaxation features, whose characteristic temperatures are not related to the crystal-field splitting energies, can be observed in a highly correlated system. The observation of such features provides weighty evidence in support of the important role of strong correlations in the compound being studied. Moreover, the calculations performed in the self-consistent approach indicate that in the case of strong relaxation broadening the upper crystal-field levels have a significant influence on the experimentally measured characteristics, which, as was previously assumed, are determined only by the properties of the ground state of a highly correlated system. An example of the properties of the ground magnetic state of ions in highly correlated systems, which can be subject to a significant influence from crystal-field excited states, is the residual width of the magnetic quasielastic neutron scattering peak at zero temperature $\Gamma_G(T=0)$. In the case of soft crystal-field splittings, employment of the widely accepted phenomenological formula $T^*=\Gamma_G(T=0)/2$ to determine the characteristic temperature T^* of the Kondo system is in need of additional analysis.

In conclusion, we wish to note that the proposed approach should be useful in the case of the analysis of systems in which the crystal-field states transform into more complicated objects as a result of strong electron correlations. Such systems include concentrated Kondo systems, in which the rare-earth ions form a coherent lattice. In this case the localized crystal-field ground-state levels transform into a coherent continuum,[41] which has been termed a spin fluid. When there are sufficiently soft crystal-field splittings, a spin fluid undergoes strong interactions with localized excited states,[42] which should produce features in the relaxation of the magnetic states. Since crystal-field states are nothing more than well defined levels, this relaxation cannot be studied in a neutron scattering experiment within the proposed method. However, the relaxation in the magnetic subsystem should have a significant influence on the spectroscopic characteristics of the system that can be detected using resonance methods, such as muon spin rotation (μSR) and nuclear magnetic resonance. These processes can be studied experimentally and calculated theoretically after the proposed formalism is appropriately generalized.

We sincerely thank P. A. Alekseev, V. N. Lazukov, and S. V. Maleev for some critical discussions. This work was supported by the Russian Fund for Fundamental Research (Project 95-02-04250a), the International Association INTAS (Project 93-2834), and the Netherlands Organization for Scientific Research NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) (Project 07-30-002).

*)E-mail: andry@kurm.polyn.kiae.su

[1)]Below we shall consider a case which corresponds to the conditions of the proposed experimental method. In this situation the concentration of the paramagnetic label is chosen small enough that the interactions between the impurities can be neglected in each specific case.[29] For this reason, the influence of the impurity on the state of the conduction electrons can be neglected.[30]

[2)]All the expressions presented above were obtained without consideration of the vertex corrections (Fig. 1d),[37] which, however, are insignificant in the case of sufficiently small dimensionless interaction constants $g_{MM'}^2 \ll 1$.

[3)]Consideration of the vertex corrections in the perturbative approach, unlike (44), leads to correction of the coefficient in front of the exponentially small natural width (41).

[1] J. P. Rice, B. G. Pazol, D. M. Ginsberg, T. J. Moran, and M. B. Weissman, J. Low Temp. Phys. **72**, 345 (1988).
[2] I. M. Lifshits and A. M. Kosevich, Zh. Eksp. Teor. Fiz. **29**, 739 (1955) [Sov. Phys. JETP **2**, 636 (1956)].
[3] E. N. Adams and T. D. Holstein, J. Phys. Chem. Solids **10**, 253 (1959).
[4] B. S. Chandrasekhar, Phys. Lett. **6**, 27 (1963).
[5] P. Thalmeier and B. Lüthi, in *Handbook on the Physics and Chemistry of Rare Earths*, K. A. Gschneider, Jr. and L. Eyring (eds.), Elsevier, Amsterdam (1991), Vol. 14, p. 225.
[6] V. L. Ginzburg and G. P. Motulevich, Usp. Fiz. Nauk **55**, 469 (1955).
[7] M. M. Sushchinsky, *Raman Spectra of Molecules and Crystals*, Israel Program for Scientific Translations, Jerusalem (1973).
[8] P. A. Aleksandrov and S. M. Yakubenya, in *Proceedings of the International Conference on the Science and Technology of Defect Control in*

Semiconductors, Yokohama, 1989, K. Sumino (ed.) (1990), Vol. **2, p.** 1605.
[9] P. A. Alekseev, I. P. Sadikov, Yu. L. Shitikov, E. M. Savitskii, O. D. Chistyakov, and J. K. Kjems, Phys. Status Solidi B **114**, 161 (1982).
[10] P. A. Alekseev, I. P. Sadikov, V. P. Kolyadin, A. V. Mirmel'shteĭn, and N. B. Kol'chugina, JETP Lett. **39**, 580 (1984).
[11] P. A. Alekseev, W. Bührer, V. N. Lazukov, E. V. Nefedova, I. P. Sadikov, O. D. Chistyakov, and M. Zolliker, Physica B **217**, 241 (1996).
[12] P. A. Alekseev, V. N. Lazukov, I. P. Sadikov, and A. Yu. Rumyantsev, JETP Lett. **43**, 758 (1986).
[13] P. A. Alekseev, V. N. Lazukov, I. P. Sadikov, and A. Yu. Rumyantsev, J. Magn. Magn. Mater. **75**, 323 (1988).
[14] P. A. Alekseev, E. S. Klement'ev, V. N. Lazukov, E. V. Nefedova, I. P. Sadikov, N. D. Efremova, L. D. Finkel'shteĭn, and N. B. Kol'chugina, Fiz. Met. Metalloved. **77**, 60 (1994).
[15] P. A. Alekseev, E. S. Klement'ev, V. N. Lazukov, I. P. Sadikov, M. N. Khlopkin, M. Adams, A. Yu. Muzychka, I. L. Sashin, N. B. Kol'chugina, and O. D. Chistyakov, JETP Lett. **63**, 1000 (1996).
[16] E. S. Clementyev, P. A. Alekseev, M. N. Khlopkin, V. N. Lazukov, I. P. Sadikov, W. Bührer, and A. Yu. Muzychka, Physica B **234–236**, 864 (1997).
[17] K. W. Becker, P. Fulde, and J. Keller, Z. Phys. B **28**, 9 (1977).
[18] K. Sugawara, Phys. Status Solidi B **92**, 317 (1979).
[19] S. V. Maleyev, Phys. Rev. B **50**, 302 (1994).
[20] L. C. Lopes and B. Coqblin, Phys. Rev. B **33**, 1804 (1986).
[21] L. C. Lopes and B. Coqblin, Phys. Rev. B **38**, 6807 (1988).
[22] J. M. Wills and B. R. Cooper, Phys. Rev. B **36**, 3809 (1987).
[23] N. Kiossus, J. M. Wills, and B. R. Cooper, J. Appl. Phys. **63**, 3683 (1988).
[24] N. Kiossus, J. M. Wills, and B. R. Cooper, Phys. Rev. B **44**, 10 003 (1991).
[25] Q. G. Sheng and B. R. Cooper, Phys. Rev. B **50**, 965 (1994).
[26] J. R. Schrieffer and P. A. Wolff, Phys. Rev. **149**, 491 (1966).
[27] B. Coqblin and J. R. Schrieffer, Phys. Rev. **185**, 847 (1969).
[28] B. Cornut and B. Coqblin, Phys. Rev. B **5**, 4541 (1972).
[29] A. A. Abrikosov, *Fundamentals of the Theory of Metals*, Elsevier, New York (1988).
[30] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, *Quantum Field Theoretical Methods in Statistical Physics*, Dover, New York (1963).
[31] B. R. Cooper, R. Siemann, D. Yang *et al.*, in *The Handbook on the Physics and Chemistry of the Actinides*, A. J. Freeman and G. H. Lander (eds.), North-Holland, Amsterdam (1985), Vol. 2 p. 435.
[32] A. S. Mishenko, JETP Lett. **66**, 487 (1997).
[33] A. A. Abrikosov, Physica (Amsterdam) **2**, 21 (1965).
[34] A. A. Abrikosov and A. A. Migdal, J. Low Temp. Phys. **3**, 519 (1970).
[35] G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **41**, 1241 (1961) [Sov. Phys. JETP **14**, 886 (1962)].
[36] G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **42**, 1658 (1962) [Sov. Phys. JETP **15**, 1151 (1962)].
[37] S. V. Maleev, Teor. Mat. Fiz. **4**, 86 (1970).
[38] S. V. Maleev, Zh. Éksp. Teor. Fiz. **79**, 1995 (1980) [Sov. Phys. JETP **52**, 1008 (1980)]; Zh. Éksp. Teor. Fiz. **84**, 260 (1983) [Sov. Phys. JETP **57**, 149 (1983)].
[39] N. Grewe and F. Steglich, in *Handbook on the Physics and Chemistry of Rare Earths*, K. A. Gschneider, Jr. and L. Eyring (eds.), Elsevier, Amsterdam (1991), Vol. 14, p. 343.
[40] P. Fulde and M. Loewenhaupt, Adv. Phys. **34**, 589 (1985).
[41] K. A. Kikoin, M. N. Kiselev, and A. S. Mishchenko, Zh. Éksp. Teor. Fiz. **112**, 729 (1997) [JETP **85**, 399 (1997)].
[42] Yu. Kagan, K. A. Kikoin, and A. S. Mishchenko, Phys. Rev. B **55**, 12 348 (1997).

Translated by P. Shelnitz

Magnetic and resonance properties of $LiCu_2O_2$ single crystals

A. M. Vorotynov, A. I. Pankrats, G. A. Petrakovskiĭ,*) and K. A. Sablina

L. V. Kirenskiĭ Institute of Physics, Siberian Branch of the Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

W. Paszkowicz and H. Szymczak

Institute of Physics, Polish Academy of Sciences, 02-668 Warsaw, Poland

(Submitted 31 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1866–1876 (May 1998)

We study the structural, magnetic, and resonance properties of $LiCu_2O_2$ single crystals grown by the spontaneous crystallization method. The data are interpreted on the assumption that the crystalline structure of the grown single crystals is orthorhombic. Long-range antiferromagnetic order sets in at temperatures below 22.5 K, while above this temperature the dependence of the magnetic susceptibility has a shape characteristic of interacting antiferromagnetic Heisenberg chains. We hypothesize that long-range magnetic order sets in below 22.5 K through the destruction of the ideal ladder structure of $LiCu_2O_2$ because of partial redistribution of copper and lithium ions at the crystal lattice sites and because of the presence of other defects in the crystalline structure. © *1998 American Institute of Physics.* [S1063-7761(98)02105-2]

1. INTRODUCTION

The discovery of high-T_c superconductivity initiated a new stage in studies of oxocuprates, which are not superconductors but contain in their crystalline structure various fragments of Cu–O characteristic of high-T_c superconductors. The oxocuprates Bi_2CuO_4 and $CuGeO_3$, which were studied earlier,[1–3] in their stoichiometric state contain only bivalent ions Cu^{2+} in quadrant and octahedral oxygen surroundings, respectively.

The present paper is a first report on a study of magnetic and resonance properties of $LiCu_2O_2$ single crystals. In contrast to Bi_2CuO_4 and $CuGeO_3$, which contain only bivalent copper ions, in the $LiCu_2O_2$ single crystal there are univalent and bivalent copper ions. The magnetic Cu^{2+} copper ions in this crystal are in pyramidal oxygen surroundings. These features determined our interest in studying the physical properties of $LiCu_2O_2$ single crystals.

2. SAMPLES AND THE EXPERIMENTAL METHOD

The $LiCu_2O_2$ single crystals were grown by slowly cooling the melt at a rate of 3–5 degrees per hour. The mixture for the melt consisted of Li_2CO_3 and CuO in a 1.2-to-1 ratio. The mixture was placed in an alundum crucible, which in turn was place inside a ZrO_2 crucible. The temperature regime was chosen with allowance for dehydration of the reagents and for the decomposition of Li_2CO_3 in the heating process. The peak temperature of the melt was 1150 °C. The time during which the melt was exposed to the peak temperature depended on the amount of mixture placed in the crucible, the size of the crucible, the size of the grains, and the extent to which the powders of the initial reagent could be mixed, and amounted to about three to four hours. The crystals were shiny black plates pccupying positions parallel to the surface of the fusing agent.

The crystals were drawn mechanically. The maximum dimensions of the crystals produced in this way were 3-by-4-by-1 mm.[3] Prolonged storage in the open air was found to lead to the formation of a mat coating on the crystal's surface.

Earlier x-ray studies of the samples[4] showed that the crystalline structure is orthorhombic with unit cell parameters $a=5.725$ Å, $b=2.857$ Å, and $c=12.409$ Å. These parameters coincide with the data of Ref. 5.

Magnetic susceptibility was measured by vibrating-reed and SQUID magnetometers in the temperature interval from 300 K to 4.2 K.

In our resonance measurements we used a magnetic-resonance spectrometer with a pulsed magnetic field with wavelength of 8 and 6 mm in the temperature interval from 300 K to 4.2 K. Magnetic resonance in the paramagnetic region was studied using an SE/X-2544 EPR spectrometer with $\lambda=3$ cm.

3. EXPERIMENTAL RESULTS

Figure 1 depicts the temperature curves for the magnetic susceptibility of $LiCu_2O_2$ measured in an 80-Oe magnetic field along the **c** axis and along two mutually perpendicular directions in the ab plane. Estimates of the diamagnetic contribution and the Van Vleck susceptibility show that both contributions are smaller than the measured values by a factor of ten.

The susceptibility along the **c** axis is higher than in the other two directions over the entire temperature range. The

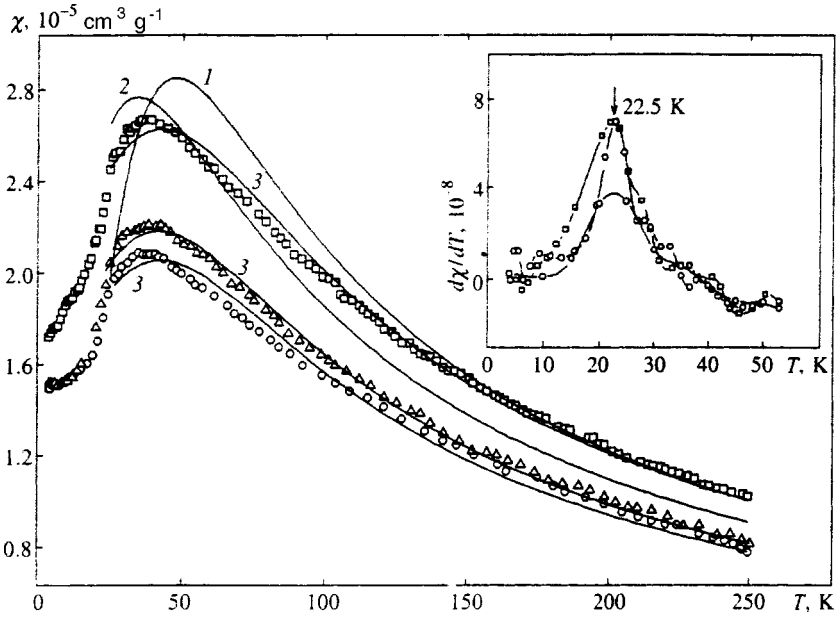

FIG. 1. Temperature curves for the magnetic susceptibility of $LiCu_2O_2$: □—**H**∥**c**, and ○ and △—**H**⊥ **c**. The theoretical curves are denoted as follows: *1*—alternatingly spaced magnetic chain, *2*—two-dimensional Heisenberg model, and *3*—one-dimensional model with interchain interaction. The inset depicts the temperature dependence of the first temperature derivatives of susceptibility.

high-temperature part of the reciprocal susceptibility is used to determine the values of the paramagnetic Néel temperatures in all three directions:

$$\Theta_c = -47\ \text{K}, \quad \Theta_{\perp 1} = -50\ \text{K}, \quad \Theta_{\perp 2} = -38\ \text{K}.$$

These values are lower than the value $\Theta = -75$ K for polycrystalline $LiCu_2O_2$ given in Ref. 6. The corresponding values of the effective magnetic moment are

$$\mu_c = 2.01\mu_B, \quad \mu_{\perp 1} = 1.83\mu_B, \quad \mu_{\perp 2} = 1.76\mu_B,$$

which are close to the theoretical value of $1.73\mu_B$ for Cu^{2+}.

Figure 2 depicts the field curves for magnetization measured at 4.2 K along the **c** axis and along two mutually perpendicular directions in the *ab* plane. All three curves represent a linear dependence up to 15 kOe. No residual magnetic moment or hysteresis phenomena were detected.

The results of studies of the temperature dependence of the EPR linewidth for three orientations of the magnetic field (along the **c** axis and in the *ab* plane) are depicted in Fig. 3. As the temperature lowers, the EPR linewidth sharply increases as T approaches $T \simeq 30$ K. Below this temperature, no EPR signal was detected with the magnetic field directed along the **c** axis. When the field was parallel to the *ab* plane, below $T \simeq 30$ K we observed resonant absorption, whose intensity was found to diminish as the temperature was lowered still further. Below 15 K this resonance signal was not detected.

The angular curves for the EPR linewidth measured at room temperature are depicted in Fig. 4. When the magnetic field rotates in the *ab* plane, the linewidth varies in sinusoidally with a period of $\pi/2$. When the magnetic field rotates in a plane perpendicular to the *ab* plane, the function $A + B(1+\cos^2\theta)$ provides a good description of the angular dependence of ΔH, with the fitting parameters A and B depending on the orientation of this plane in relation to the crystallographic axes in the *ab* plane.

The temperature curves for EPR fields with both field orientations are depicted in Fig. 5. Figure 6 depicts the angular curves for the g factors measured at room temperature. For **H**∥**c** we have $g_c = 2.225$, while in the *ab* plane the g factor slowly varies sinusoidally with a period of $\pi/2$ and extremal values 2.00 and 1.95.

Figure 7 depicts the frequency vs. field dependence of magnetic resonance, measured at $T = 4.2$ K in **H**∥**c**. The dashed straight line shows the linear dependence $\nu = \gamma H$, where γ corresponds to the value g_c measured at room tem-

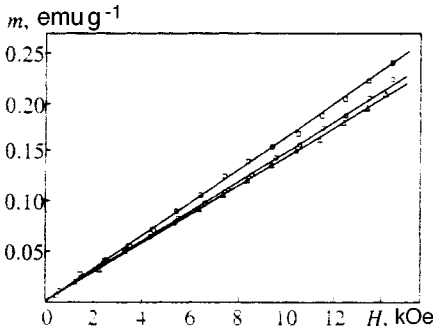

FIG. 2. Field curves for the magnetization of $LiCu_2O_2$ at $T = 4.2$ K: □—**H**∥**c**, and ○ and △—**H**⊥ **c**.

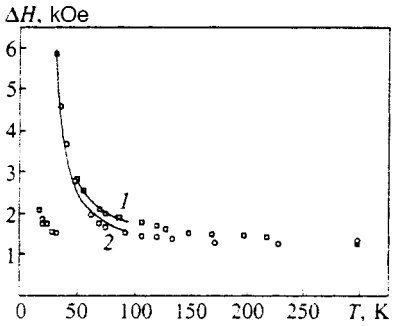

FIG. 3. Temperature curves for the EPR linewidth in $LiCu_2O_2$ at $\nu = 9.4$ GHz: *1*—field **H** is parallel to the *ab* plane, *2*—**H**∥**c**. The solid curves represent the results of power-law calculations.

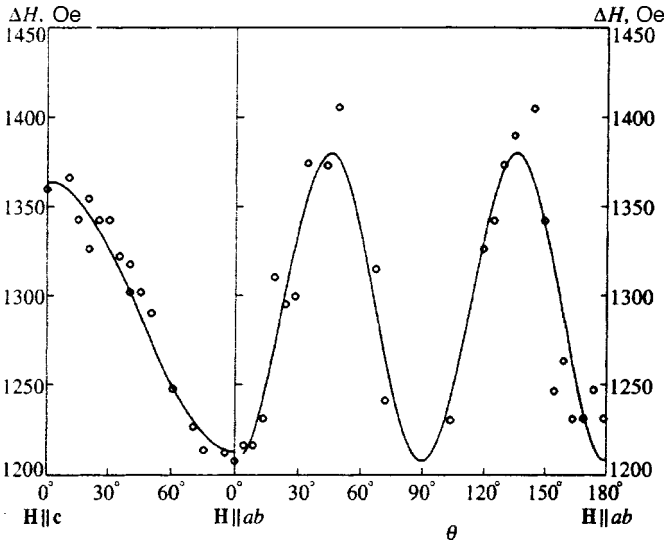

FIG. 4. Angular curves for the EPR linewidth in $LiCu_2O_2$ at room temperature at $\nu = 9.4$ GHz.

perature. The frequency vs. field dependence has a gap and is described by the polynomial

$$\nu = \nu_c + aH + bH^2, \tag{1}$$

where $\nu_c = 29.83$ GHz, $a = 0.5754$ GHz kOe^{-1}, and $b = 0.0265$ GHz kOe^{-2}. Note that the frequency vs. field dependence does not asymptotically approach the linear dependence $\nu = \gamma H$ as the field becomes stronger; it intersects this linear dependence at $H \simeq 13$ kOe.

We measured the temperature behavior of the resonance fields in $\mathbf{H} \| \mathbf{c}$ at two frequencies, 44.61 and 46.27 GHz. Under the assumption that the coefficients a and b in Eq. (1) are temperature-independent, we calculated the temperature dependence of ν_c (Fig. 8). According to this dependence, the ν_c (T) curve is approximately equal to zero at $T \simeq 23$ K.

Figure 9 depicts the results of measuring the angular curves for the resonance field when the magnetic field is rotated in two mutually perpendicular planes containing the **c** axis. Within experimental error, both curves coincide. For the magnetic field in the ab plane the resonance field sharply increases, which made it impossible to measure the frequency vs. field dependence for this orientation.

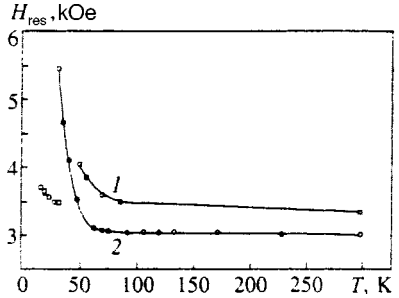

FIG. 5. Temperature curves for the EPR field in $LiCu_2O_2$ at $\nu = 9.4$ GHz: *1*—$\mathbf{H} \| ab$, and *2*—$\mathbf{H} \| \mathbf{c}$.

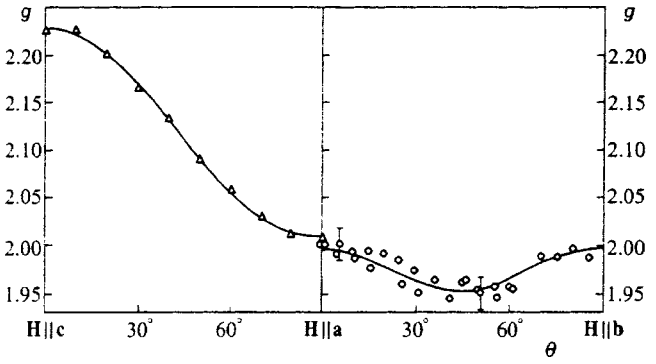

FIG. 6. Angular curves for the g factor of $LiCu_2O_2$ at room temperature and $\nu = 9.4$ GHz.

4. DISCUSSION

The crystalline structure of $LiCu_2O_2$ was determined by Hibble *et al.*[6] by the x-ray method. They interpreted the structure $LiCu_2O_2$ as tetragonal with the space group $P4_2/nmc$ and $a = 5.714$ Åand $c = 12.401$ Å. Later, however, Berger *et al.*[5] interpreted the crystalline structure of $LiCu_2O_2$ as orthorhombic with the space group $Pnma$ and the unit cell parameters $a = 5.72$ Å, $b = 2.86$ Å, and $c = 12.4$ Å. They explain why Hibble *et al.*[6] inferred tetrahedral symmetry in $LiCu_2O_2$ by pointing out that $a = 2b$, so that there is crystal twinning, and this causes the x-ray spectra to have quasitetragonal symmetry. In Ref. 6 it was assumed that depending on the rate of melt cooling and other conditions of synthesis, the crystal symmetry may be either tetragonal or orthorhombic. Thus, the difference in determining the space group of the $LiCu_2O_2$ crystal may be due to the different origin of the samples.

The crystalline structure of $LiCu_2O_2$ with orthorhombic symmetry is depicted in Fig. 10a. The Cu^{2+} ions are at the base of the pyramid consisting of oxygen ions and are connected into chains along the **b** axisof the crystal. The magnetic structure of $LiCu_2O_2$ in this case is formed by two exchange-coupled chains of Cu^{2+} ions oriented along the **b** axis of the crystal and belonging to two neighboring atomic planes that are perpendicular to the **c** axis of the crystal. These planes form a layer in which the adjacent exchange-coupled pairs of chains consisting of copper ions are far from

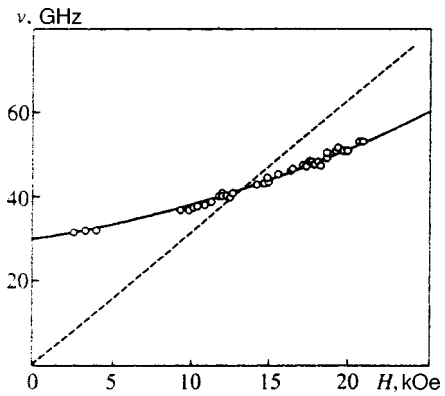

FIG. 7. Frequency vs. field dependence of the AFMR signal in $LiCu_2O_2$ at $T = 4.2$ K in $\mathbf{H} \| \mathbf{c}$. The solid curve represents the theoretical curve corresponding to Eq. (1).

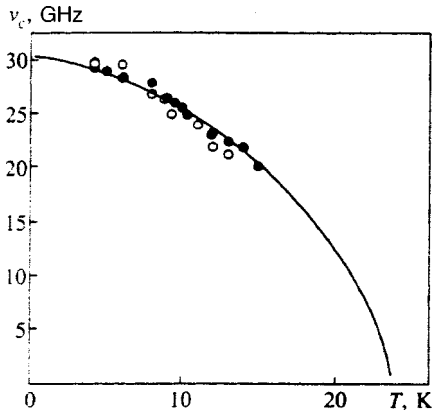

FIG. 8. Temperature curve for the AFMR gap in $LiCu_2O_2$: ○—ν = 44.61 GHz, and ●—ν = 46.27 GHz.

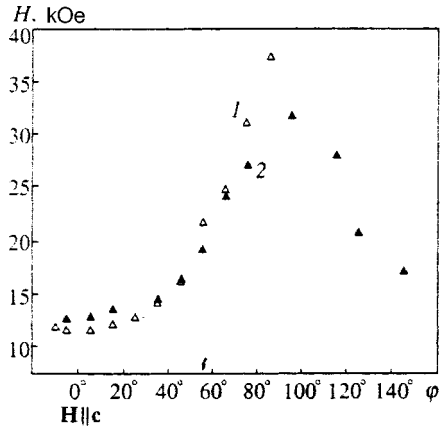

FIG. 9. Angular curves for the resonance field of the AFMR signal in $LiCu_2O_2$ at $T = 4.2$ K and $\nu = 41.62$ GHz. The data *1* and *2* were obtained as a result of rotating the sample in two mutually perpendicular planes containing the **c** axis of the crystal.

each other and are separated by chains consisting of lithium ions also oriented along the **b** axis. Each layer is separated from a neighboring layer of the same kind by a plane containing nonmagnetic Cu^+ ions. Thus, we have described a magnetic structure consisting of isolated pairs of exchange-coupled chains of Cu^{2+} ions, with a fragment shown in Fig. 10b.

Such magnetic structures are known in the literature as ladder systems,[7] and lately there has been an upsurge of interest in these structures. A ladder system consisting of two chains, or a two-leg ladder, is described by two exchange interactions, the intrachain interaction and the interchain interaction. In our case, as Fig. 10b clearly shows, we must introduce three exchange interactions to describe the magnetic structure. A similar structure was observed in the $KCuCl_3$ crystal by Tanaka *et al.*,[8] who found that such a magnetic structure is a ladder system.

If there is twinning of the orthorhombic crystal, the crystal becomes divided into domains in such a way that in neighboring domains the crystallographic **b** axes are turned through 90° in relation to each other. However, in this case too the magnetic structure within a single domain is a ladder one. Since in crystal twinning the domains are usually macroscopic, the magnetic structure of the crystal as a whole can also be considered a ladder system.

When the symmetry is tetragonal, the crystalline structure of $LiCu_2O_2$ proposed by Hibble *et al.*[6] differs from the orthorhombic in that in two neighboring atomic planes the ion copper chains are perpendicular to each other. In this case the magnetic structure is not a ladder one; more likely, it forms a two-dimensional system that is a bulk ''lattice'' of exchange-coupled chains consisting of neighboring atomic planes.

Thus, depending on the type of structure (orthorhombic or tetragonal) and the strength of the exchange coupling between the chains, $LiCu_2O_2$ can be in the form of the following magnetic structures: noninteracting magnetic chains, exchange-coupled pairs of magnetic chains, or a two-dimensional magnetic structure consisting of two neighboring atomic planes with the chains in these planes perpendicular to each other.

Since x-ray studies of our samples of $LiCu_2O_2$ confirmed the orthorhombic symmetry of the crystal, below we discuss the various magnetic structures that can arise in this case.

As shown in Refs. 7–10, the ground state of a ladder system with $S = 1/2$ is nonmagnetic and is characterized by an energy gap between the ground and excited magnetic states. The presence of such a gap leads to a situation in which the magnetic susceptibility drops exponentially to zero as the temperature falls below a certain critical value.

The temperature curves for the susceptibility depicted in Fig. 1 do not exhibit the low-temperature exponential drop characteristic of ladder systems. More than that, the peaks in

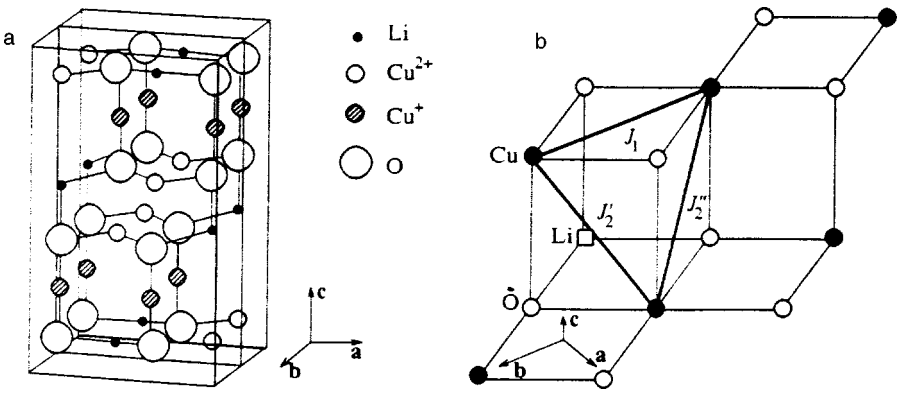

FIG. 10. (a) The crystalline structure of $LiCu_2O_2$ with orthorhombic symmetry (according to the data of Ref. 5). (b) A fragment of the crystalline structure of $LiCu_2O_2$ illustrating exchange interactions.

the first temperature derivatives for all three direction of the magnetic field at $T=22.5$ K (the inset in Fig. 1) suggest that long-range magnetic order sets in below this temperature, which is assumed to be the Néel temperature T_N. Presumably, the existence of long-range magnetic order is also supported by the presence of a gap in the magnetic resonance spectrum at temperatures below 22.5 K (Figs. 7 and 8). The presence of magnetic order in the crystal suggests that there is no need to examine in detail the case of noninteracting magnetic chains.

Strictly speaking, the absence of an exponential drop in susceptibility makes it impossible to reject with certainty the ladder nature of the magnetic structure of $LiCu_2O_2$. Fukuyama *et al.*[11] studied the ladder system $SrCu_2O_3$ and found that zinc ion impurities and, probably, other structural defects can give rise to a magnetic phase and, at sufficiently high concentrations of such defects, to the formation of a Néel state. According to Watanabe *et al.*,[10] strong anisotropy of the exchange interaction in a chain can also lead to such a state.

The presence of a broad maximum in the temperature curve for the susceptibility and the inequality $T_N/T_{\chi\max}=0.59<1$ indicate that the magnetic structure of $LiCu_2O_2$ is low-dimensional, with the antiferromagnetic interaction in a chain realized through a 90° bond. This situation is encountered in oxocuprates and, in particular, in $CuGeO_3$. It was given a theoretical basis by Geertsma and Khomskii.[12]

Figure 1 depicts the theoretical temperature curves for the susceptibility calculated in the two- and one-dimensional Bonner–Fisher models for planes and interacting antiferromagnetic Heisenberg chains.[13,14] Figure 1 also depicts the theoretical temperature curve for the magnetic susceptibility of an alternatingly spaced magnetic chain, whose behavior is similar to that of a ladder system.[15] The best agreement with experiment is achieved for a system of interacting antiferromagnetic chains with $J_1=-31.5$ K and $|J_2|=0.06$ K with the external magnetic field **H** parallel to the **c** axis (here J_1 and J_2 are the intra- and interchain exchange integrals). The value of the intrachain exchange interaction J_1, estimated from the relation $T_{\chi\max}=1.32|J_1|S(S+1)$ (see Ref. 16), is $|J_1|=39.6$ K.

A comparison of the theoretical and experimental curves suggests that the magnetic structure in $LiCu_2O_2$ is quasi-one-dimensional. The absence of an exponential decay in susceptibility to zero (a decay characteristic of ladder systems) in this case is probably due to the presence of defects in the sample, which destroy the purely ladder state.

The crystalline structure of $LiCu_2O_2$ shows that the Li^+ and Cu^{2+} ions have the same pyramidal oxygen surroundings. This probably facilitates the situation in which in some of the lithium and bivalent copper ions change places. The possibility of such redistribution is corroborated, for instance, by the fact that doping $CuMg_2O_3$ with lithium ions leads to a uniform distribution of the impurity ions among the Cu and Mg positions,[17] while in $LiCu_3O_3$ the Li and Cu ions are distributed among equivalent positions statistically.[6]

On the one hand, such redistribution of ions violates the homogeneity of the ladder structure but, on the other, an exchange coupling develops between isolated ladder pairs of chains via the Cu^{2+} ions that have occupied the positions of lithium ions. When the number of such copper ions is sufficiently large, long-range magnetic order of the Néel type may set in at a finite temperature in the crystal,[11] and the temperature curve for magnetic susceptibility is described by the theoretical curve for interacting chains. Here by exchange interaction between chains we mean an average value of the exchange interactions between chains within a two-leg ladder and the interaction between ladder pairs via the copper ions introduced into the lithium chains.

Another fact that speaks in favor of partial redistribution of copper and lithium ions in $LiCu_2O_2$ is that after hardening the samples from high temperatures, the resonance and magnetic properties undergo a considerable change (as shown by preliminary measurement). Here we have not excluded the possibility that the redistribution of cations may lead to formation of a tetragonal phase. The problem of how heat treatment affects the structural and magnetic properties of $LiCu_2O_2$ requires a special study.

We cannot exclude the possibility that conduction electrons also participate in the formation of long-range magnetic order in $LiCu_2O_2$. Most likely, the presence of copper ions with different valencies leads to a situation in which the electrical resistivity of $LiCu_2O_2$ at room temperature is roughly 10^2–$10^3\,\Omega$cm, which is smaller by several orders of magnitude than in $CuGeO_3$ or Bi_2CuO_4.

The frequency vs. field dependence of magnetic resonance at low temperatures and, in particular, the presence of a gap in the magnetic excitation spectrum (Fig. 7) and the temperature dependence of this gap (Fig. 8) also suggest that antiferromagnetic order sets in at temperatures lower than 22.5 K. However, the gap in the AFMR spectrum is extremely narrow for orthorhombic antiferromagnets. As is well known,[18] the size of the gap in the AFMR spectrum for an orthorhombic antiferromagnet is given by the following expression:

$$\nu_c=\gamma\sqrt{2H_EH_a'},$$

where H_E is the exchange field, and, depending on the orientation of the external magnetic field with respect to the crystallographic axis, H_a' is one of the two effective fields H_{a1} and H_{a2} describing the anisotropy of a biaxial antiferromagnet or the difference of these fields. A possible situation (at least in principle) is when H_{a1} and H_{a2} are close in value and their difference determines a small value of the energy gap. But in this case, too, the absence of significant anomalies in the field curves for magnetization in a field $H_c=\sqrt{2H_EH_a'}\approx 11.8$ kOe due to a spin-flop transition remains unexplained.

On the other hand, the frequency vs. field dependence does not approach the linear dependence $\nu=\gamma H$ asymptotically as the magnetic field strength grows, as was to be expected in orthorhombic antiferromagnets, but intersects it. A frequency vs. field dependence of this kind can be observed, for instance, in antiferromagnets with a triangular (noncollinear) magnetic structure.[19]

When the external magnetic field was at right angles to the **c** axis of the crystal, we were unable to measure the

frequency vs. field dependence at frequencies above 37 GHz because of a sharp increase in the resonance field. However, on the basis of the angular dependence of the resonance field of AFMR (Fig. 9) it can be assumed that the external magnetic field deviates from the **c** axis of the crystal, the gap in the AFMR spectrum diminishes and the slope of the frequency vs. field curve decreases.

It is quite possible that all the unusual properties of AFMR are due to the formation of a noncollinear magnetic structure for $T<T_N$. Such a structure may develop because of frustration of the exchange interaction in the triangular bond configuration (see Fig. 10b).

The increase in the EPR linewidth with the temperature decreasing below ≈ 80 K (Fig. 3) is described by the theoretical formula $\Delta H \propto [(T-T_N)/T_N]^{-n}$. The best agreement with the experimental results is achieved at $n=1.28$ and $n=1.35$ for $\mathbf{H}\|\mathbf{c}$ and $\mathbf{H}\perp\mathbf{c}$, respectively. These values of the critical indices are close to the values $n=1.1-1.2$ determined from experiments with the quasi-one-dimensional magnetic materials $CuCl_2\cdot 2NC_5H_5$, $CsNiCl_3$, and $RbNiCl_3$ (see Refs. 20 and 21).

An analysis of linewidths and the values of g factors measured at room temperature (Figs. 4 and 6) suggests the following. The large value of ΔH cannot be explained by dipole–dipole and effective exchange ($J_{\text{eff}} \propto 3\Theta_N/2zS(S+1)$ interactions and is determined by the anisotropic exchange interaction. When the magnetic field changes its orientation from the **c** axis to the ab plane, ΔH and the g-factor exhibit a typical angular dependence, $\propto A(1+\cos^2\theta)$. However, in the ab plane both are characterized by 90° periodicity, while in orthorhombic crystals this angular dependence has a 180° period and tetragonal crystals exhibit no such angular dependence, provided that we ignore contributions of higher-order exchange interactions. We believe that such discrepancy is due to crystal twinning, in which the crystallographic axes in neighboring domains are rotated about the **c** axis through an angle of 90°.

5. CONCLUSION

Our study of the structural, magnetostatic, and resonance properties of $LiCu_2O_2$ single crystals leads to the following conclusions.

The $LiCu_2O_2$ compound is a quasi-low-dimensional magnetic material. At $T=22.5$ K, long-range antiferromagnetic order sets in in a $LiCu_2O_2$ single crystal. It is hypothesized that the magnetic structure in the magnetically ordered phase is noncollinear.

Our experimental results have been interpreted on the basis of the assumption that the crystalline structure of the samples is orthorhombic (the space group $Pnma$). In an ideal crystalline structure of $LiCu_2O_2$ the positions of the atoms predetermine a magnetic structure of a two-leg ladder. However, defects in the crystalline structure, among which the most probable are the partial redistribution of copper and lithium ions Cu^{2+} and Li^{+} in chains and the oxygen nonstoichiometry, destroy the singlet state (a state characteristic of ladder systems with spin $S=1/2$) and introduce long-range magnetic order in the system.

The authors are grateful to K. S. Aleksandrov and O. A. Bayukov for useful discussions in the course of writing the paper, to D. A. Velikanov for measuring magnetic susceptibility, and to N. I. Kiselev for measuring electrical resistivity. This work was made possible by a grant from the Krasnoyarsk Regional Scientific Fund (Grant No. 6F0004).

*)E-mail: gap@post.krascience.rssi.ru

[1] G. A. Petrakovskii, K. A. Sablina, A. M. Vorotinov, V. N. Vasiliev, A. I. Kruglik, A. D. Balaev, D. A. Velikanov, and N. I. Kiselev, Solid State Commun. **79**, 317 (1991).
[2] A. I. Pankrats, G. A. Petrakovskii, and K. A. Sablina, Solid State Commun. **91**, 121 (1994).
[3] G. A. Petrakovskiĭ, K. A. Sablina, A. M. Vorotynov, A. I. Kruglik, A. G. Klimenko, A. D. Balayev, and S. S. Aplesnin, Zh. Éksp. Teor. Fiz. **98**, 1382 (1990) [Sov. Phys. JETP **71**, 772 (1990)].
[4] W. Paszkowicz, A. Vorotinov, K. Sablina, and G. Petrakovskii, in *Abs. Int. Conf. on X-Ray Powder Diffraction Analysis of Real Structure of Matter*, Slovakia, 1995, p. 68
[5] R. Berger, P. Onnerud, and R. Tellgren, J. Alloys Compd. **184**, 315 (1992).
[6] S. J. Hibble, J. Kohler, and A. Simon, J. Solid State Chem. **88**, 534 (1990).
[7] E. Dagotto, J. Riera, and D. Scalapino, Phys. Rev. B **45**, 5744 (1992).
[8] H. Tanaka, K. Takatsu, W. Shiramura, and T. Ono, J. Phys. Soc. Jpn. **65**, 1945 (1996).
[9] M. Azuma, Z. Hiroi, M. Takano, K. Ishida and Y. Kitaoka, Phys. Rev. Lett. **73**, 3463 (1994).
[10] H. Watanabe, K. Nomura, and S. Takada, J. Phys. Soc. Jpn. **62**, 2845 (1993).
[11] H. Fukuyama, N. Nagaosa, M. Saito, and T. Tamimoto, J. Phys. Soc. Jpn. **65**, 2377 (1996).
[12] W. Geertsma and D. Khomskii, Phys. Rev. B **54**, 3011 (1996).
[13] J. Bonner and M. E. Fisher, Phys. Rev. **135**, A640 (1964).
[14] G. S. Rushbrook and P. J. Wood, Mol. Phys. **1**, 257 (1958).
[15] W. E. Hatfield, J. Appl. Phys. **52**, 1985 (1981).
[16] T. de Neef, Phys. Rev. B **13**, 4141 (1976).
[17] W. Winkelmann, H. A. Graf, and N. H. Anderson, Phys. Rev. B **49**, 310 (1994).
[18] E. A. Turov, *Physical Properties of Magnetically Ordered Crystals* [in Russian], Publishing House of the USSR Academy of Sciences, Moscow (1963), p. 165.
[19] L. A. Prozorova, V. I. Marchenko, and Yu. V. Krasnyak, JETP Lett. **41**, 637 (1985).
[20] Y. Ajiro, H. Matsukawa, and Y. Shuni-Ichi, Phys. Lett. A **72**, 367 (1979).
[21] V. V. Velichko and E. A. Petrakovskaia, in *Proc. 1st Int. Conf. on Physics of Magnetic Materials*, Poland, 1980.

Translated by Eugene Yankovsky

Quantum oscillations of resistance and magnetization in the degenerate semiconductor *n*-$HgCr_2Se_4$

A. D. Balaev and V. A. Gavrichkov*)

L. V. Kirenskiĭ Institute of Physics, Siberian Branch of the Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

S. G. Ovchinnikov

L. V. Kirenskiĭ Institute of Physics, Siberian Branch of the Russian Academy of Sciences, 660036 Krasnoyarsk, Russia; Krasnoyarsk State University, 660074 Krasnoyarsk, Russia

V. K. Chernov

Krasnoyarsk State University, 660074 Krasnoyarsk, Russia

T. G. Aminov and G. G. Shabunina

N. S. Kurnakov Institute of General and Inorganic Chemistry, Russian Academy of Sciences, 117907 Moscow, Russia

(Submitted 3 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 1877–1882 (May 1998)

In the magnetic field range $\Delta H = 8-60$ kOe we observed and studied the anomalous oscillations in the magnetic field dependence of the resistance and magnetization of single crystals of *n*-$HgCr_2Se_4$. The absence of periodicity in $1/H$ in the $\Delta H = 8-20$ kOe range can be explained by the non-Fermi-liquid behavior of the electron subsystem and agrees with the theory of the de Haas–van Alphen in systems with intermediate valence. In stronger fields, $\Delta H = 20-60$ kOe, the amplitude of the fundamental harmonic decreases, with the number and amplitude of the higher-order harmonics increasing. As a result, noise is superimposed on the signal as magnetic field strength grows. The temperature dependence of the magnetization is the sum of the monotonic spin-wave contribution and the oscillating part. © *1998 American Institute of Physics.* [S1063-7761(98)02205-7]

1. INTRODUCTION

Landau oscillations, which arise because of quantization of electron orbits in a magnetic field, usually lead to oscillations (periodic in $1/H$) of the thermodynamic and kinetic characteristics of degenerate Fermi systems. If the dependence of the chemical potential μ on temperature and magnetic field differs from that in standard Fermi liquids, e.g., due to strong electron correlations, anomalous quantum oscillations may occur. In an earlier paper[1] we reported the detection of temperature oscillations of magnetization in single crystals of the degenerate magnetic semiconductor *n*-$HgCr_2Se_4$. In the present paper we give the results of observations of weakly periodic and aperiodic (in $1/H$) oscillations of the resistance R and magnetization M of the same samples. While observing temperature oscillations of magnetization required complicated multiparameter processing of experimental data,[1] oscillations in the field dependence of M and R are visible without any theoretical processing.

The ferromagnetic semiconductor *n*-$HgCr_2Se_4$ has a temperature-independent electron concentration n of roughly 10^{18}cm^{-3} in the temperature range $4.2 < T < 77$ K (in which the investigations took place) and a high carrier mobility $\sim 10^3\text{cm}^2\ \text{V}^{-1}\text{s}$ at 77 K, which makes it possible to observe Landau oscillations in fields up to 10 kOe. Non-Fermi-liquid

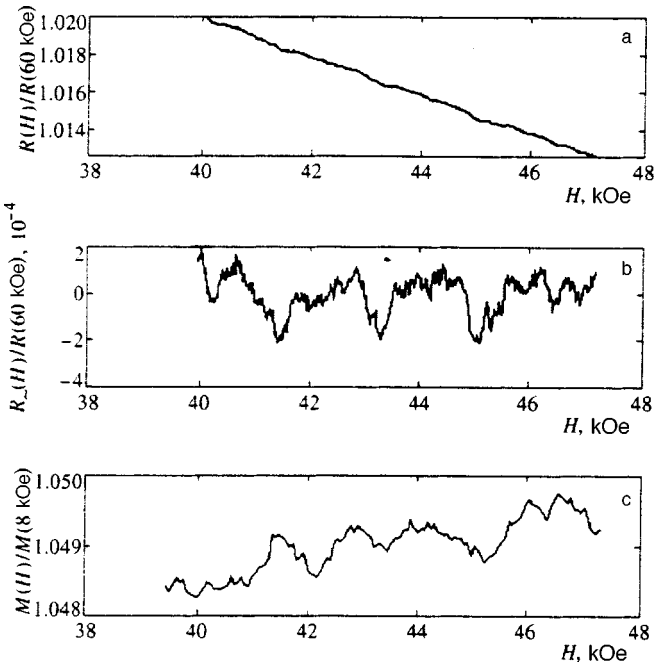

FIG. 1. Magnetic field dependence of the total resistance (a), the oscillating part of the resistance (b), and the relative magnetization (c) of the degenerate semiconductor *n*-$HgCr_2Se_4$ measured at $T = 4.2$ K.

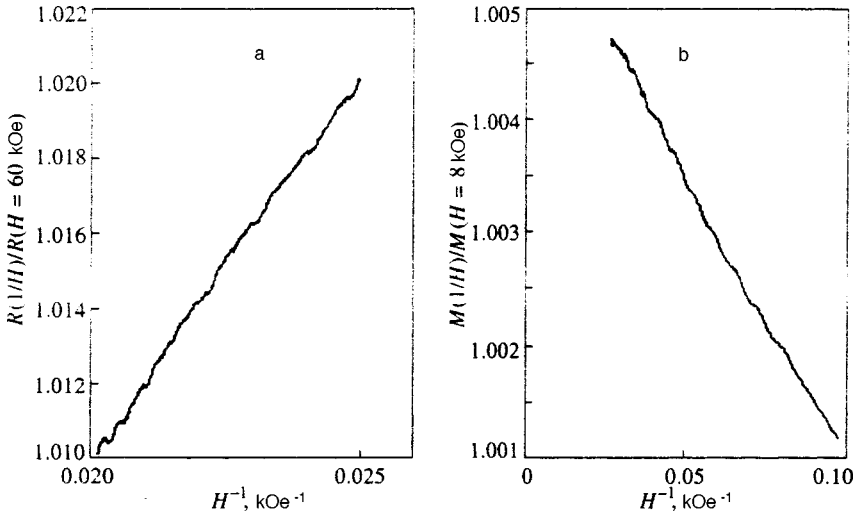

FIG. 2. Dependence of resistance (a) and magnetization (b) on the reciprocal magnetic field for n-$HgCr_2Se_4$ at $T=4.2$ K.

effects in $HgCr_2Se_4$ are due to the presence of a localized d-level of chromium Ω near the bottom of the conduction band.[2] In a degenerate n-type semiconductor, the chemical potential μ is pinned near the localized level and is weakly dependent on T and H, while the bottom of the conduction band is shifted in proportion to the magnetization $M(T,H)$. As a result the function $\mu(T,H)$ acquires non-Fermi-liquid corrections. Temperature oscillations of magnetization in such a multielectron model were predicted in Ref. 3, and after the necessary experimental work was done (see Ref. 1), a more comprehensive theory of the de Haas–van Alphen effect was developed in a recent paper by Val'kov and Dzebisashvili.[4] In the temperature interval $\mu_B \ll T \ll \hbar\omega_c$, realized because of the small effective carrier mass $m \sim 0.01 m_e$, the new theory yields a temperature and magnetic dependence of magnetization such that the chemical potential measured from the bottom of the conduction band can be written as[4]

$$\mu(T,H)=\mu(0)-\frac{J}{2}Z\left(\frac{3}{2}\right)t^{3/2}+Jt\sqrt{h}-\frac{35}{96\pi}J\sqrt{th}, \qquad (1)$$

where $Z(3/2)=2.612$, J is the $3d$-exchange integral, $t=T/4\pi IS$, and $h=\mu_B H/IS$, with I the parameter of exchange between the neighboring spins of a d-ion, and S the spin of the d-ion. Since the oscillating part of the thermodynamic potential and its derivatives is determined by the factor

$$\sin\left(2\pi k\frac{\tilde{\mu}}{\hbar\omega_c}\right)=\sin\left\{\frac{2\pi k}{\hbar\omega_c}\left[\mu(0)-\frac{J}{2}Z\left(\frac{3}{2}\right)t^{3/2}+Jt\sqrt{h}-\frac{35}{96\pi}J\sqrt{th}\right]\right\}, \qquad (2)$$

it is clear that the field dependence of the oscillation phase assumes the form

$$\varphi\sim\frac{a}{H}+\frac{b}{\sqrt{H}},$$

i.e., the periodicity of oscillations in $1/H$ is violated, although the oscillations remain.

2. FIELD DEPENDENCE OF MAGNETIZATION AND RESISTANCE

The magnetization of samples was measured by an automatic vibrating-reed magnetometer with a superconducting solenoid in fields up to 60 kOe (Ref. 5) at $T=4.2$ K. The magnetic field was applied along the $\langle 100\rangle$ axis of the crystal. Longitudinal magnetoresistance was measured by the four-contact method.

Figure 1 depicts the oscillating part $R_\sim$ of the resistance and the magnetization M as functions of the magnetic field strength, while Fig. 2 depicts the dependence of the resistance R on the reciprocal field strength. From Fig. 1a we see that a magnetoresistance linear in the field strength is superimposed on the oscillating part of the resistance, so we represent $R(H)$ in the form

$$R(H)=R_\sim(H)-cH, \qquad (3)$$

where $c=R(60\ \text{kOe})\times 10^{-4}\text{kOe}^{-1}$ is a parameter. The oscillating part $R_\sim(H)$ defined in this manner is depicted in Fig. 1b. A comparison of the $R_\sim$ vs. H curve in Fig. 1b and the M vs. H curve in Fig. 1c shows that the extrema in the two curves $R_\sim(H)$ and $M(H)$ coincide. As expected, the oscillations are periodic neither in H nor in $1/H$. The spectral density of the signal depicted in Fig. 2b has a smeared peak corresponding to an approximate period in $1/H$ in 0.8×10^{-6}Oe. However, it is possible to detect a distinct signal only in the magnetic field range $\Delta H=8$–20 kOe (Fig. 3a). In stronger fields, e.g., in the range $\Delta H=20$–60 kOe, the amplitude of the fundamental harmonic decreases, with the number and amplitude of the higher-order harmonics increasing. As a result, noise is superimposed on the signal (Fig. 3b). Knowing the period, we calculated the area S of the extremal cross section of the Fermi surface and the carrier concentration n. The results were $S=9.3\times 10^{13}\text{cm}^{-2}$ and $n=4.3\times 10^{18}\text{cm}^{-3}$.

The relative amplitudes of the oscillations are moderate, $\sim 10^{-4}$, but they are larger than the magnetization measurement errors by a factor of approximately ten.[5] The smallness of the magnetization oscillation amplitude can be explained by the smallness of the carrier concentration, since at n

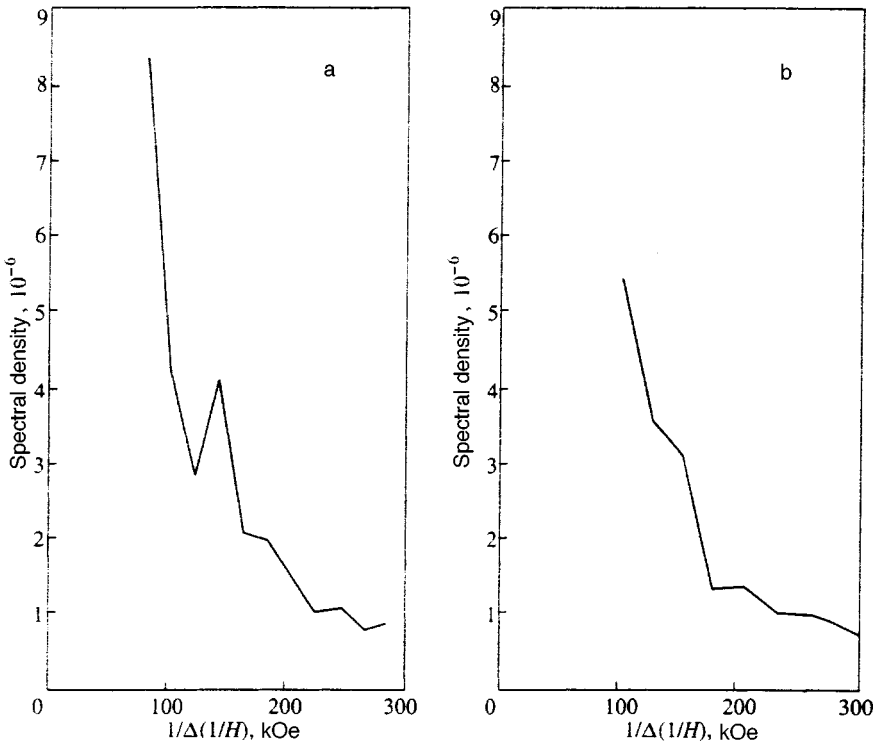

FIG. 3. Spectral density of magnetization $M(1/H)$ in magnetic field ranges $\Delta H = 8-20$ kOe (a) and 20–60 kOe (b).

$\sim 10^{18}\text{cm}^{-3}$ the number of electrons per cell is roughly 10^{-4}. The relative amplitude of resistance oscillations is approximately five times larger than the magnetization oscillation amplitude.

3. TEMPERATURE DEPENDENCE OF MAGNETIZATION AND RESISTANCE

Formula (2) shows that temperature variations cause large shifts in the chemical potential and intersections with the Landau levels, which gives rise to oscillations in the temperature dependence of the magnetization and resistance. Since the oscillation phase is a nonlinear function of temperature, the oscillations are aperiodic in T. Here, however, the damping of the oscillation amplitude with increasing temperature makes observation of a large number of temperature oscillations difficult.

The other fact that sets temperature oscillations apart from field oscillations is that temperature oscillations are masked by a complicated temperature dependence of both magnetization and resistance. Hence to identify the oscillation contribution one must subtract the monotonic parts. For instance, for the average spin the spin-wave theory yields[6]

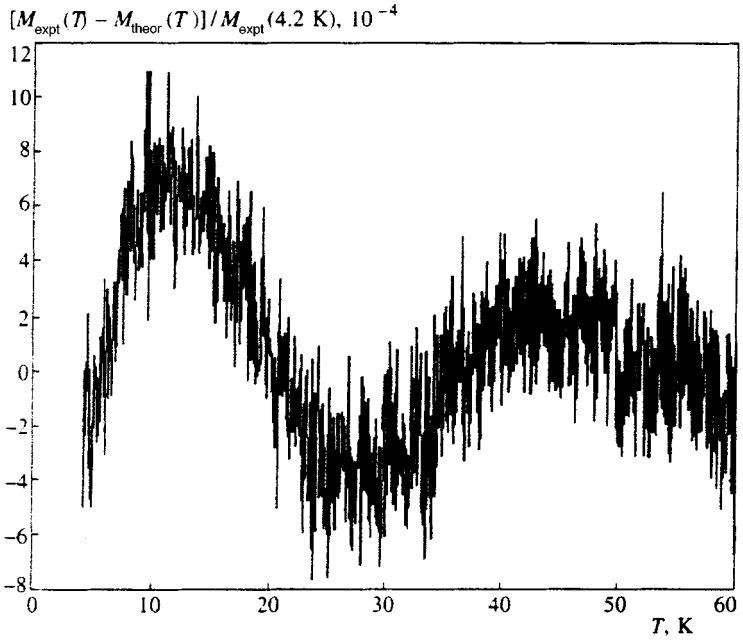

FIG. 4. Difference of the experimental curve $M_{expt}(T)$ for n-$HgCr_2Se_4$ and the theoretical curve $M_{theor}(T)$ expressed in the spin-wave approximation as a function of temperature for a field of 60 kOe.

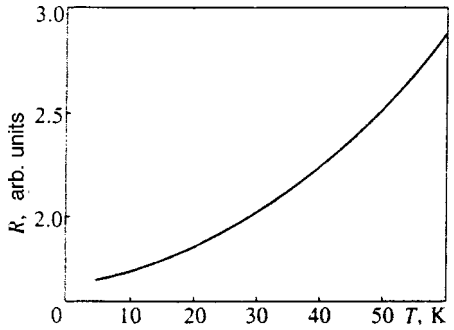

FIG. 5. Temperature dependence of the resistance of n-$HgCr_2Se_4$ in a magnetic field of 60 kOe.

$$\langle S^z\rangle_{SW}(T)=\frac{3}{2}-at^{3/2}Z_{3/2}(x)-bt^{5/2}Z_{5/2}(x)-ct^{7/2}Z_{7/2}(x), \tag{4}$$

where $x=2\mu_B H/kT$, and $Z_p(x)$ is the generalized Riemann zeta function,

$$Z_p(x)=\sum_{n=1}^{\infty}\frac{\exp(-nx)}{n^p}.$$

The parameters a, b, and c were found by fitting the results to the experimental curve $M_{\text{expt}}(T)$ measured in a field $H=60$ kOe via the simplex method. The values are $a=0.8499$, $d=-0.5545$, and $c=0.1294$.

The difference of the measured curve $M_{\text{expt}}(T)$ and the theoretical curve $M_{\text{theor}}(T)$ determined via (4) is depicted in Fig. 4. Thus, the total temperature dependence of the magnetization can be written as the sum of the monotonic curve (4) and the oscillating part.

The temperature dependence of the electrical resistance measured in $H=60$ kOe is depicted in Fig. 5. As in the case of the $M_{\text{expt}}(T)$ curve, to identify the oscillations we must subtract the monotonic temperature dependence, which is not related to Landau quantization. Although the various mechanisms of the temperature dependence of the electrical resistance of magnetic semiconductors have been thoroughly studied,[7] quantitative comparison with experimental results requires special calculations that allow for the behavior of the band structure of n-$HgCr_2Se_3$. Such calculations are outside the scope of the present paper.

5. CONCLUSION

In this paper we have shown that the Shubnikov–de Haas and de Haas–van Alphen effects in the degenerate semiconductor n-$HgCr_2Se_3$ can be observed in magnetic fields of the 8–60 kOe range but that their field dependence is not described by functions periodic in $1/H$, in contrast to the case of an ordinary Fermi liquid. The non-Fermi-liquid nature of the temperature dependence of the chemical potential also gives rise to quantum temperature oscillations of magnetization.

This work was made possible by a grant from the Russian Fund for Fundamental Research (Grant No. 96-02-16075).

*)E-mail: gav@iph.krasnoyarsk.su

[1] S. G. Ovchinnikov, V. K. Chernov, A. D. Balaev, N. B. Ivanova, B. P. Khrustalev, and V. A. Levshin, JETP Lett. **64**, 642 (1995).
[2] V. A. Gavrichkov, M. Sh. Erukhimov, and S. G. Ovchinnikov, Fiz. Tverd. Tela (Leningrad) **29**, 527 (1987) [Sov. Phys. Solid State **29**, 298 (1987)].
[3] V. V. Val'kov and S. G. Ovchinnikov, Fiz. Tverd. Tela (Leningrad) **24**, 1801 (1982) [Sov. Phys. Solid State **24**, 1024 (1982)].
[4] V. V. Val'kov and D. Dzebisashvili, Zh. Éksp. Teor. Fiz. **111**, 654 (1997) [JETP **84**, 360 (1997)].
[5] A. D. Balaev, Yu. B. Boyarshinov, M. M. Karpenko, and B. P. Khrustalev, Prib. Tekh. Eksp. No. **3**, 167 (1985).
[6] S. V. Vonsovskii, *Magnetism*, 2 vols., Wiley, New York (1974).
[7] É. L. Nagaev, *The Physics of Magnetic Semiconductors*, Nauka, Moscow (1979); transl. Mir, Moscow (1983).

Translated by Eugene Yankovsky

Raman scattering of light and IR absorption in carbon nanotubes

A. V. Bazhenov,*) V. V. Kveder, A. A. Maksimov, I. I. Tartakovskiĭ, R. A. Oganyan, Yu. A. Ossipyan, and A. I. Shalynin

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

(Submitted 17 December 1997)
Zh. Éksp. Teor. Fiz. **113**, 1883–1891 (May 1998)

Raman light scattering and IR absorption spectra of samples containing multilayer carbon nanotubes in different stages of purification by the selective oxidation technique have been investigated. It was found that the Raman spectra of carbon nanotubes exhibit softening of the mode at 1582 cm^{-1} corresponding to E_{2g} vibrations of graphite hexagons and a line at 120 cm^{-1} due to the radial vibrations of nanotubes. In IR absorption spectra measured in the region of 0.07–0.3 eV, several sets of lines with a spacing of 15 meV (120 cm^{-1}) between lines of each group have been detected. We suggest that each group corresponds to electron transitions generating electron–hole pairs in semiconducting nanotubes and contains a phononless 00-line and its phonon replicas with spacing between them equal to the ''breathing'' mode energy of 120 cm^{-1}. Measurements of electric conductivity at a frequency of 9300 MHz indicate that, in addition to semiconducting nanotubes, the samples contain nanotubes with properties of a highly disordered semimetal.
[S1063-7761(98)02305-1]

1. INTRODUCTION

It is widely known that the standard arc-discharge technique, which is used in fabrication of fullerenes from graphite, allows one to produce macroscopic quantities of multilayered carbon nanotubes[1–3] with an outside diameter of about 10 nm and a length of several microns to tens of microns. X-ray and electron-microscope experiments[4,5] show that these nanotubes consist of concentric cylindrical graphite layers imbedded one inside another (this configuration is traditionally called the ''Russian-doll'' model). Depending on the regime of the arc-discharge synthesis, nanotubes can have various numbers of layers, different chiral angles, and different diameters.

It is obvious that carbon nanotubes can have peculiar electronic properties and constitute a new class of quasi-one-dimensional structures. Since their radius is very small, one can expect manifestations of size quantization of their electron wave functions in the direction perpendicular to the tube diameter, which should give rise to one-dimensional energy bands in their electron spectra. According to theoretical calculations, the electron spectrum of a nanotube can be either metallic or semiconducting, depending on its configuration.[6–12] Unfortunately, published experimental data concerning these novel species are still sparse because of considerable technical problems. Measurements of macroscopic samples containing large fractions of nanotubes indicate that the electrical conductivity is not metallic and drops with decreasing temperature (see, for example, Refs. 13 and 14). The conductivity as a function of temperature can be described in terms of incoherent, thermally activated tunneling. In these experiments, however, the sample resistance is a function of resistivities of both nanotubes and contacts between them. With the help of modern lithographic nanotechniques, the dc conductivity of several isolated nanotubes was measured.[15,16] The electric contacts were narrow metallic strips deposited in vacuum across the nanotubes laid on insulating substrates. These measurements indicate that the shapes of nanotube conductivity curves plotted against temperature vary considerably: several nanotubes demonstrated thermally activated (semiconducting) conductivity with activation energies of about 0.1 and 0.3 eV, whereas the conductivity of others behaved like that of a semimetal or a semiconductor with a very narrow band gap.

It is obvious that additional research using various experimental techniques is needed for better understanding of electronic properties of nanotubes. This paper reports measurements of IR absorption and optical Raman scattering performed on either isolated strands of carbon nanotubes or samples composed of nanotubes isolated from one another, which were purified of inclusions of graphite particles and amorphous carbon by the selective oxidation technique.

2. FABRICATION OF SAMPLES AND MEASUREMENT TECHNIQUES

Carbon nanotubes were fabricated using the standard technique in a dc electric arc discharge between two electrodes from pure graphite in helium atmosphere. The helium pressure was 600 mbar, the arc length was about 2 mm at a constant discharge current of 80 A. The central part of the carbon deposit generated by the arc discharge, which contains a higher proportion of nanotubes, was separated by drilling. The resulting material was composed of needle-shaped pieces (Fig. 1a), which were ''bundles'' of nanotubes baked to one another. In order to reduce the content of small graphite particles and soot, the samples were annealed for 30

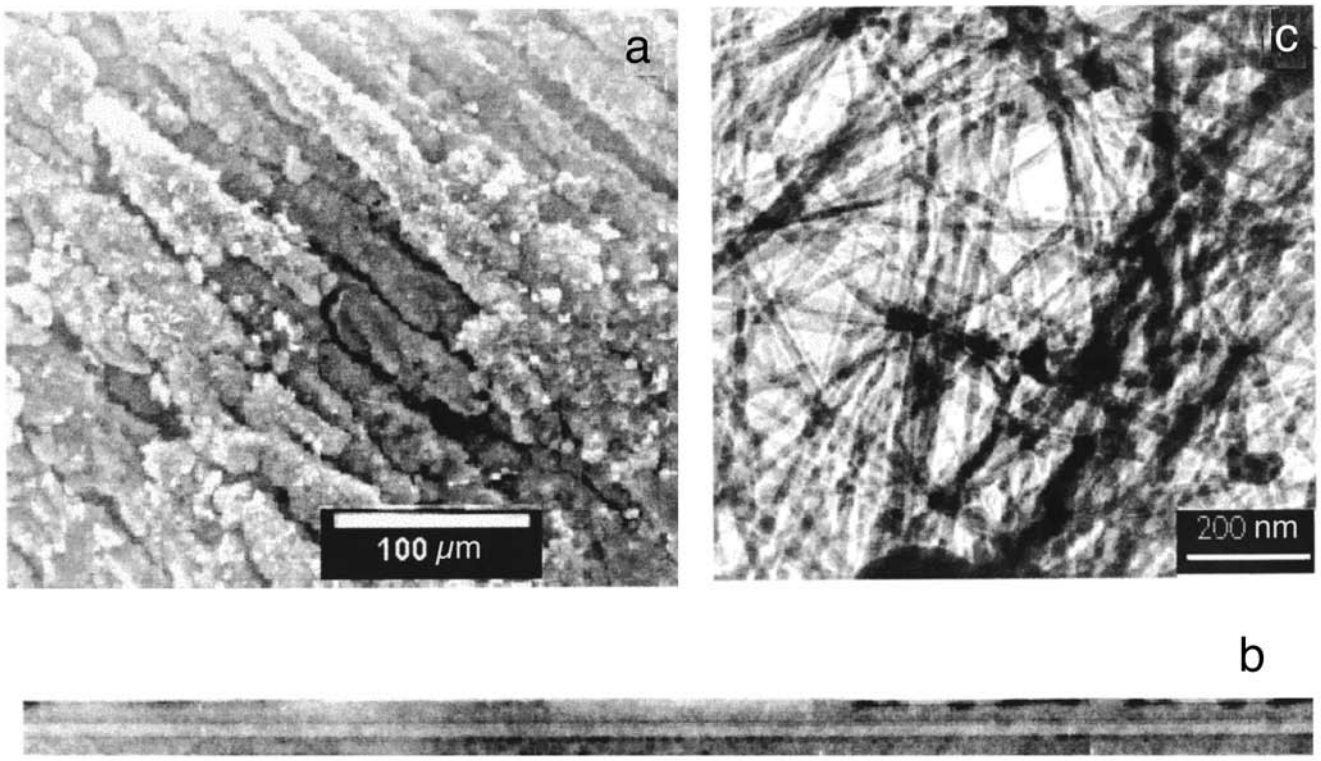

FIG. 1. (a) Image of needle-shaped conglomerates of carbon nanotubes in type-*1* samples in a scanning tunnel microscope; (b, c) typical images of a type-*3* sample in a transmitting tunnel microscope. The sample has the shape of ''cotton wool'' formed of isolated carbon nanotubes (c), one of them is shown on a larger scale (b).

min in air at a temperature of 650 °C. The fraction of weight lost in the process was about 10%. Hereafter the samples manufactured by this method are called type *1*. They were composed largely from nanotube bundles and small carbon particles baked together in needle-shaped pieces. The next step in purification of nanotube samples was partial destruction of these pieces by ultrasound in ethanol and a second annealing of the produced powder in air at a temperature of 670–700 °C until about 80% of the sample weight was burnt out (type-*2* samples) or until 97–99% of weight loss (type-*3* samples). Figure 1c shows an electron microscope image of a type-*3* sample. Such samples had the form of ''cotton wool'' composed of isolated pure carbon nanotubes. Figure 1b shows as an illustration one of such nanotubes imaged at a large magnification factor. Typical outside diameters of nanotubes in our samples were 10–15 nm. According to electron microscope measurements, type-*2* samples were like type 3 samples, but contained a larger proportion of soot particles and other undesired carbon species.

Raman spectra were measured on a Dilor XY-500 spectrometer equipped with a multichannel optical detector and Olympus BH-2 optical microscope, which allowed us to measure Raman spectra with a spatial resolution of less than 1 μm. Samples were excited by an Ar^+ laser at a wavelength $\lambda = 488$ nm. Measurements were performed at room temperature.

IR absorption spectra were recorded between 0.05 and 0.6 eV using a Bruker IFS-113v Fourier-transform spectrometer at room temperature.

In addition to optical measurements, we also measured microwave absorption of samples at a frequency of 9300 MHz at temperatures between 1.5 and 300 K. In order to get rid of the skin-effect, samples for microwave experiments were fabricated by mixing nanotube powder with melted paraffin in the ultrasound field. After solidification of the paraffin, a ball-shaped sample with a volume of about 5 mm^3 and containing 0.5 mg of nanotubes was cut out. The sample was placed in a rectangular microwave cavity, and the real and imaginary parts of its permittivity, $\varepsilon = \varepsilon' + i\varepsilon''$, were derived from measurements of the cavity resonant frequency F_r and FWHM ΔF of its resonant curve.

3. RESULTS AND DISCUSSION

Figure 2 shows Raman spectra of carbon nanotube samples of types *1*, *2*, and *3* in the region of high-frequency vibrations. For comparison, a Raman spectrum of pyrolithic graphite measured under the same conditions is given. The crystalline graphite spectrum contains an isolated high-frequency line, which corresponds, as is well known, to the Raman-active E_{2g} mode due to vibrations in carbon hexagons with an energy of 1582 cm^{-1}. It is also well known that, if the translation symmetry breaking is essential, for example, if the graphite particle size is smaller than 100 nm, another line with an energy of 1360 cm^{-1} can be seen in the Raman spectra. This line is due to a vibrational mode, which is usually Raman-silent, so the ratio R between the intensities of the 1360 cm^{-1} line and the allowed line at 1582 cm^{-1}

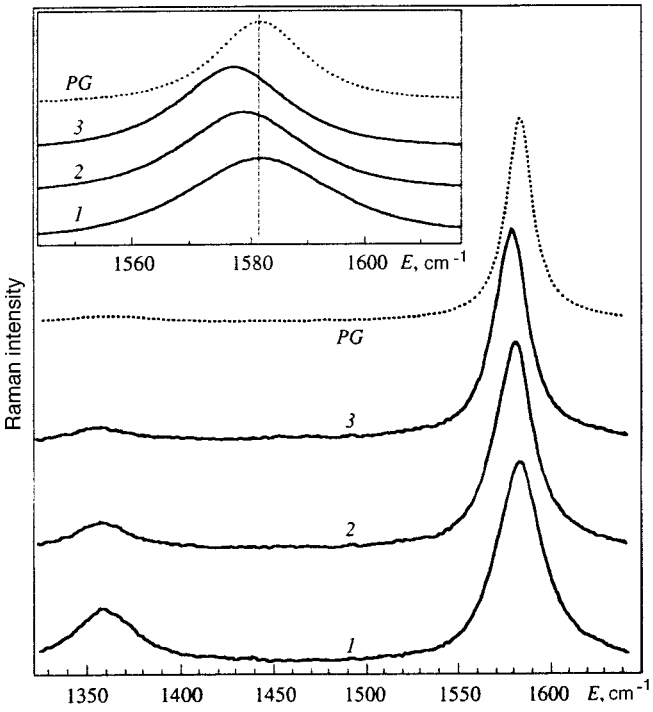

FIG. 2. Raman spectra for samples of type *1*, *2*, and *3* in the region of high-frequency vibrations. For comparison, the spectrum of pyrolithic graphite is also shown (dashed curve). The inset shows on a larger scale the spectral lines corresponding to the E_{2g} mode in samples *2* and *3* in comparison with the graphite line.

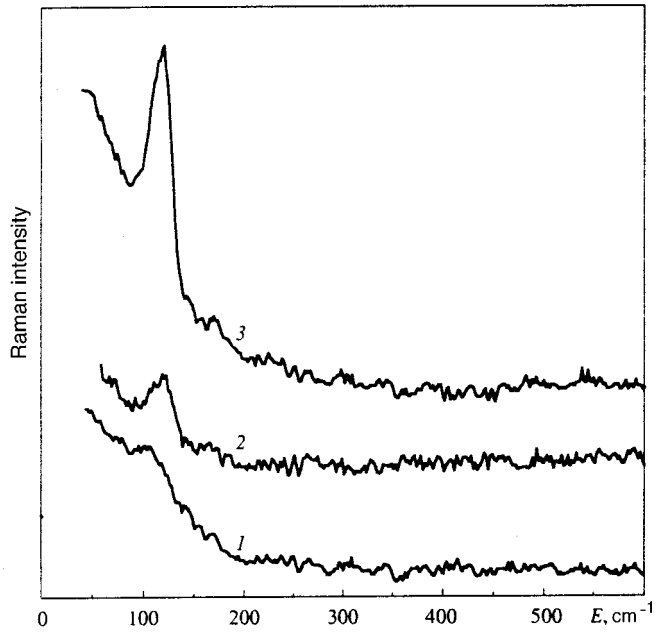

FIG. 3. Low-frequency sections of Raman spectra of samples of type *1*, *2*, and *3*.

is commonly used as a criterion of the presence of very small graphite particles in a sample and in estimating their size.[17] If the particle size L is larger than 100 nm, R is less than 0.03, whereas at $L=10$ nm this ratio is about 0.4.[17]

One can see in Fig. 2 that the Raman line in the energy region near 1360 cm^{-1} in spectra of samples of type *1* and *2* has a fairly large amplitude, which indicates the presence of very small fragments of graphite layers. At first sight, this line could be attributed to nanotubes. In fact, given the graphite layer dimension close to the nanotube diameter of 10 to 15 nm, one could expect $R=I_{1360}/I_{1582}$ to be 0.2–0.4,[17] which is in fair agreement with measurements of the type-*1* spectrum (curve *1* in Fig. 2). But the amplitude of the 1360 cm^{-1} line drops with the sample is purified by selective oxidation, and R is also smaller in the type-*3* sample, which contains a small proportion of impurities, than in the initial type-*1* material. This means that the parameter R is very small for nanotubes, and the forbidden 1360 cm^{-1} line cannot be used as a test of the presence of nanotubes. It seems that in type-*1* and type-*2* samples this line is due to the presence of a large amount of very small nanoparticles of graphite and amorphous carbon.

Figure 2 also shows that the main difference between the spectra of the pure nanotube sample (curve *3*) and that of graphite in this energy region is the shift of the E_{2g} mode from 1582 cm^{-1}, which is typical of usual plane graphite layers, to 1575 cm^{-1}. This shift is in good agreement with calculations,[18] which indicate that bending a graphite layer into the shape of a cylinder with a radius approximately equal to that of nanotubes should soften this mode by 8–15 cm^{-1}. Thus, this shift of the Raman line to the low-frequency region detected in the type-*3* sample indicates that the main contributors to this spectrum are nanotubes.

Figure 3 shows Raman spectra of samples of type *1*, *2*, and *3* in the region of small Raman shifts. One can see a line corresponding to a vibrational energy of about 120 cm^{-1} in the Raman spectrum of the nanotube sample. This line is absent in samples of pyrolithic graphite, and its intensity increases as nanotube samples are purified of undesired species. Thus, the low-frequency mode in the region about 120 cm^{-1} is a feature of nanotubes and, apparently, corresponds to the ''breathing'' mode of oscillations in cylindrical nanotube walls.[19,20]

Figure 4 shows optical absorption spectra, $K(E)=-\log(T(E))$, where $T(E)$ is the optical transmission spectrum of a sample. Curve *1* from the type-*1* sample was taken from a single needle-shaped cluster (see Fig. 1a) with dimensions of about $0.04\times0.04\times0.12$ mm^3 with the help of an IR microscope. The light spot on the sample was smaller than the sample itself, which allowed us to measure the spectrum using the traditional Fourier transform technique. One can see a considerable decrease in the absorption in the energy region below 0.15 eV, which means a lower combined density of electron states at smaller energies.

Curve *3* was taken from a type-*3* sample, which had the form of a piece of ''cotton-wool'' formed by pure nanotubes. In this specific case, measurements were performed on a thin layer of nanotubes mechanically deposited on a transparent substrate. The absorption spectrum of a similar sample recorded using a parallel light beam incident normally on the substrate resembles the spectrum of the type-*1* sample, but it has additional features in the form of oscillations on the absorption curve. Since the sample was very inhomogeneous in the latter case, it produced a great amount of diffusely scat-

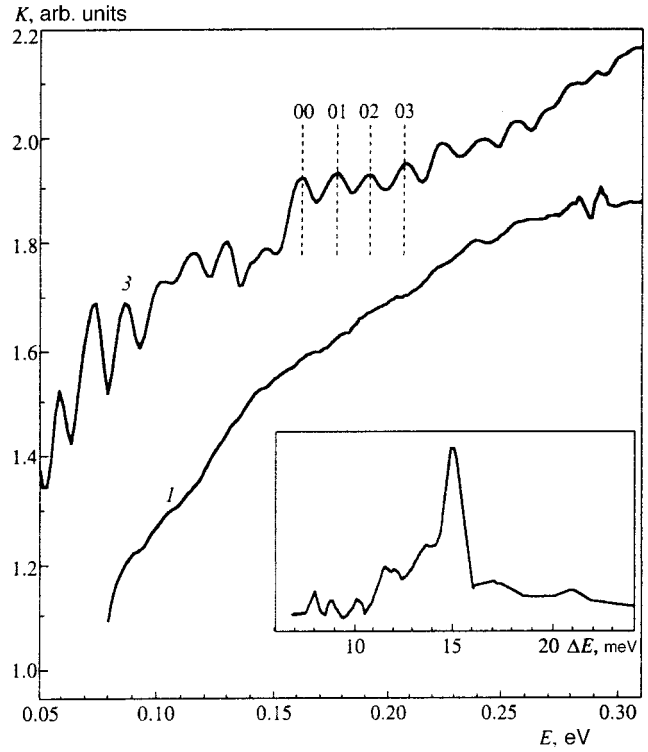

FIG. 4. IR absorption spectra at $T=300$ K of type *1* (needle-shaped nanotube clusters) and type *3* (nanotubes purified by selective oxidation) samples. The inset shows a Fourier transform of curve *3*.

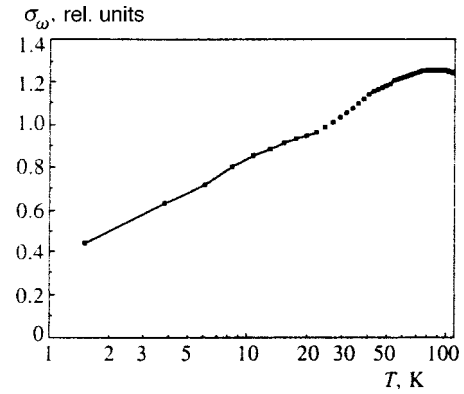

FIG. 5. Microwave electric conductivity of sample *3* versus logarithm of temperature.

tered light, and we were wary of mistaking an artefact due to light interference for a real physical effect. In order to check whether the recorded oscillations were due to interference, we calculated absorption spectra using measurements of both transmitted and diffusely scattered light collected from a large solid angle. In this case, interference effect should have been suppressed by a considerable factor. This is the spectrum shown in Fig. 4 by curve *3*. It turned out that the new oscillatory features in the 0.1–0.3-eV region were not eliminated as a result of such data processing, but on the contrary became more pronounced.

The narrow lines recorded in the 0.1–0.3-eV region can be divided into several groups, each of which includes three of four equidistant lines with a spacing of 15 meV between them. The inset in Fig. 4 shows as an illustration a Fourier transform of curve *3*. It contains a pronounced peak corresponding to the 15-meV period. This energy is consistent with the Raman peak at 120 cm^{-1} corresponding to radial vibrations of nanotube walls.

We suggest the following interpretation of the narrow lines in the absorption spectrum. Suppose that our samples contain a certain proportion of semiconducting nanotubes with a fully occupied valence band and a vacant conduction band (at $T=0$, of course), which are separated by a band gap of width E_g. The latter should, naturally, be a function of the nanotube configuration, namely its diameter, number of layers, and chiral angle. But since all these parameters are integer (or, at least, discrete), the set of possible E_g is limited. Since the density of states in a one-dimensional band is proportional to $E^{-1/2}$, where the energy is measured with respect to the band bottom, there should be a singularity in the density of states at $E=E_g$. Therefore the optical absorption spectrum of a semiconducting nanotube should have a narrow peak at E_g corresponding to phononless 00-transitions. Owing to the electron–phonon interaction, additional peaks due to phonon-assisted transitions involving one (01) or several (02, 03, etc.) phonons can be detected. In the context of this model, we can attribute the absorption spectrum shown by curve *3* to semiconducting nanotubes with different gap widths, each of which generates a phononless peak and several phonon replicas due to interaction with the 120-cm^{-1} vibrational mode seen in Raman spectra. Since two or three phonon replicas of the 00-transition are clearly seen in the spectrum, we assume that the electron–phonon interaction involving the radial mode at 120 cm^{-1} is very strong.

On the base of this model, we conclude that our samples contain notable fractions of semiconducting nanotubes with $E_g=0.07$, 0.11, 0.16, and 0.22 eV. These conclusion is consistent with measurements of the conductivity activation energy.[16,21] Specifically, activation energies of 0.1 and 0.3 eV were measured in two different nanotubes,[16] and a set of activation energies of 0.04, 0.11, and 0.23 eV were detected in electric conductivity of thin nanotube films grown on an insulating substrate.[21] If we assume that the Fermi level of a semiconducting nanotube is approximately in the middle of the band gap at low temperatures, the reported activation energies approximately correspond to gap widths $E_g=0.08$, 0.22, 0.46, and 0.6 eV.

If most of the nanotubes in our samples are semiconducting, we cannot account for the broad featureless background in the IR absorption spectra. This component of the absorption spectrum can be ascribed to the presence of highly conductive metallic or semimetallic tubes, in addition to semiconducting species. In order to verify this assumption, we performed contactless measurements of electric conductivity of our samples by the microwave absorption technique. Measurements of the type *3* sample are given in Fig. 5. It turned out that, even at $T=1.5$ K, the average nanotube conductivity σ_ω, which was derived from measurements[22,23] of $\varepsilon=\varepsilon'+i\varepsilon''$ and the number of nanotubes in the sample, was about 100–300 $\Omega^{-1}\cdot cm^{-1}$, which indicates an appreciable quantity of metallic or semimetallic nanotubes. As shown in Fig. 5, the conductivity of these highly conducting

nanotubes is roughly proportional to ln T in the temperature interval 1.5–50 K. This result is in agreement with the data by Langer *et al.*,[15] who derived a logarithmic temperature dependence of the conductivity from direct measurements of an isolated nanotube with four metallic contacts fabricated by the lithographic technique. They interpreted this logarithmic dependence in terms of weak localization due to the large contribution of quantum interference corrections to the sample resistance. This contribution drops with temperature owing to the decrease in the coherence length. Thus, in addition to semiconducting nanotubes, our samples contain a large amount of nanotubes whose conductance behaves like that of a metal with a large concentration of defects (or a semiconductor with a vanishing band gap).

The work was financially supported by INTAS (Grant 94-0157).

*)E-mail: bazhenov@issp.as.ru

[1] S. Ijima, Nature **354**, 56 (1991).
[2] T. W. Ebbesen and P. M. Ajayan, Nature **358**, 220 (1992).
[3] T. W. Ebbesen, P. M. Ajayan, H. Hiura, and K. Tanigaki, Nature **367**, 519 (1994).
[4] X. F. Zhang, X. B. Zhang, G. Van Tendeloo, S. Amelinkx, M. Op de Beeck, and J. Van Landuyt, J. Cryst. Growth **130**, 368 (1993).
[5] D. Reznik, C. H. Olk, D. A. Neumann, and J. R. D. Copley, Phys. Rev. B **52**, 116 (1995).
[6] M. S. Dresselhaus, Nature **358**, 195 (1992).
[7] J. W. Mintmire, B. I. Dunlap, and C. T. White, Phys. Rev. Lett. **68**, 631 (1992).
[8] R. Saito, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, Appl. Phys. Lett. **60**, 2204 (1992).
[9] X. Blase, L. X. Benedict, E. L. Shirley, and S. G. Louie, Phys. Rev. Lett. **72**, 1878 (1994).
[10] H. Yorikawa and S. Muramatsu, Phys. Rev. B **50**, 12203 (1994).
[11] R. Saito, G. Dresselhause, and M. S. Dresselhaus, Phys. Rev. B **50**, 14698 (1994).
[12] H. Yorikawa and S. Muramatsu, Phys. Rev. B **52**, 2723 (1995).
[13] M. Terrones, J. P. Hare, K. Hsu, H. W. Kroto, A. Lappas, W. K. Maser, A. J. Pierik, K. Prassides, R. Taylor, and D. R. M. Walton, Electrochemical Society Proceedings, **95-10**, 599 (1995).
[14] Walt A. de Heer, W. S. Bacsa, A. Chatelain, T. Gerfin, R. Humphrey-Baker, L. Forro, and D. Ugarte, Science **268**, 845 (1995).
[15] L. Langer, V. Bayot, E. Grivei, J.-P. Issi, J. P. Heremans, C. H. Olk, L. Stockman, C. Van Haesendonck, and Y. Bruynseraede, Phys. Rev. Lett. **76**, 479 (1996).
[16] T. W. Ebbesen, H. J. Lezec, H. Hiura, J. W. Bennett, H. F. Chaemi, and T. Thio, Nature (London) **382**, 54 (1996).
[17] T. C. Chieu, M. S. Dresselhaus, and M. Endo, Phys. Rev. B **26**, 10530 (1982).
[18] W. S. Bacsa, D. Ugarte, A. Chvtelain, and W. A. de Heer, Phys. Rev. B **50**, 15473 (1994).
[19] P. V. Huong, R. Cavagnat, P. M. Ajayan, and O. Stephan, Phys. Rev. B **51**, 10048 (1995).
[20] H. Jantoljak, U. Kuhlmann, C. Thomsen, S. Curran, and S. Roth, in *Proceedings of the 3rd Internat. Workshop ''Fullerenes and Atomic Clusters (IWFAC'97),''* St. Petersburg, July 1997.
[21] A. P. Korovin, T. G. Kolesnikova, T. C. Kitichenco, V. A. Vdovenkov, Z. Ya. Kosakovskaya, and V. G. Mokerov, in *Proceedings of the 3rd Internat. Workshop ''Fullerenes and Atomic Clusters (IWFAC'97),''* St. Petersburg, July 1997.
[22] L. K. H. Van Beek, Prog. Dielectr. **7**, 69 (1967).
[23] V. V. Kveder, R. Labusch, and Yu. A. Ossipyan, Phys. Status Solidi A **92**, 293 (1985).

Translation provided by the Russian Editorial office.

Optical solitons and quasisolitons

V. E. Zakharov*) and E. A. Kuznetsov†)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia
(Submitted 13 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 1892–1914 (May 1998)

Optical solitons and quasisolitons are investigated in reference to Cherenkov radiation. It is shown that both solitons and quasisolitons can exist, if the linear operator specifying their asymptotic behavior at infinity is sign-definite. In particular, the application of this criterion to stationary optical solitons shifts the soliton carrier frequency at which the first derivative of the dielectric constant with respect to the frequency vanishes. At that point the phase and group velocities coincide. Solitons and quasisolitons are absent, if the third-order dispersion is taken into account. The stability of a soliton is proved for fourth order dispersion using the sign-definiteness of the operator and integral estimates of the Sobolev type. This proof is based on the boundedness of the Hamiltonian for a fixed value of the pulse energy. © *1998 American Institute of Physics.* [S1063-7761(98)02405-6]

1. INTRODUCTION

Solitons in nonlinear optical fibers have been very popular objects of investigation since the early nineteen seventies, i.e., since the structural stability of the solitons for the Korteweg-de Vries (KdV) equation[1] and the nonlinear Schrödinger equation[2] was demonstrated and since Hasegawa and Tappert[3] subsequently proposed the use of optical solitons as data bits in fiber communications. The interest in optical solitons has increased dramatically in the last decade due to the practical achievements from the use of solitons in modern optical communication systems.[4,5] However, despite the great practical significance of optical solitons, the theory for them is far from complete.

When reference is made to optical solitons, it is assumed that their spectrum is concentrated within a certain transparency window, where the linear damping is small and dispersion effects dominate. The width of the soliton spectrum $\delta\omega$ is assumed to be fairly small compared with the frequency band $\Delta\omega$ of that window, i.e., $\delta\omega \ll \Delta\omega$. In real systems, however, the band $\Delta\omega$ is always narrower than the mean frequency of the window $\overline{\omega}$, i.e., $\Delta\omega \ll \overline{\omega}$. Thus, we have the following hierarchy of inverse characteristic times:

$$\delta\omega \ll \Delta\omega \ll \overline{\omega}. \tag{1.1}$$

These criteria permit consideration of the slow ($\tau^{-1} \sim \delta\omega$) dynamics of soliton propagation in terms of amplitude envelopes. In particular, to derive a nonlinear Schrödinger equation (NLSE), i.e., the basic model for describing optical envelope solitons, the wave number is approximated by a quadratic polynomial

$$\delta k = \frac{1}{v_{\rm gr}}\delta\omega - \frac{1}{2}\frac{\omega''}{v_{\rm gr}^3}(\delta\omega)^2. \tag{1.2}$$

Here $\delta k = k - k_0$, $\delta\omega = \omega - \omega_0$, $v_{\rm gr} = \partial\omega/\partial k$ is the group velocity, and k_0 and ω_0 are the wave number and frequency of the soliton carrier wave. However, in the frequency interval $\Delta\omega$ the dispersion of the wave can differ significantly from the quadratic approximation (1.2), although it remains small in the sense of the criterion (1.1). It is noteworthy that the existing experimental possibilities (see, for example, Ref. 6) make it possible to obtain very short pulses, for which $\delta\omega/\omega_0 < 1$. On the other hand, the efficiency of optical fibers as media for transmitting information is inversely proportional to the soliton width. Thus, practical considerations call for reducing the soliton width as much as possible.

In this paper we show that the properties of ''short'' and ''long'' solitons can be very different. For short solitons the expansion (1.2) is largely incorrect and should be replaced by the more general formula

$$\delta k - \frac{1}{v_{\rm gr}}\delta\omega = -F(\delta\omega). \tag{1.3}$$

Here $F(\zeta)$ is a certain function, which should be taken from a microscopic treatment or extracted from experimental data. Although $F(\zeta)$ can be far from the parabolic dependence (1.2), averaging over the fast time $1/\omega_0$ can be performed, providing a description of slow soliton dynamics by means of a generalized nonlinear Schrödinger equation (GNLSE). This averaging also leads to the appearance of an additional integral of motion, viz., an adiabatic invariant, which has the meaning of the pulse energy. Accordingly, owing to this invariant, the GNLSE allows a soliton solution for the envelope of the electromagnetic field $E(x,t)$ in the form a propagating pulse with the additional phase multiplier $e^{i\lambda x}$:

$$E(x, t - x/v_{\rm gr}) = e^{i\lambda x}\psi(t - x/v_{\rm gr} + \beta x), \quad v_{\rm gr}^{-1} \gg \beta.$$

The main result of this paper is as follows. Solitons can exist, if $L(\zeta) = \lambda - \beta\zeta + F(\zeta)$ is a positive (or negative) definite function for all ζ. This criterion is the basic selection rule for solitons. If this criterion is not satisfied, the soliton loses its energy through Cherenkov radiation and ceases to exist after a certain time. This occurs, for example, if $F(\zeta)$ is a third-degree polynomial.

Even if $L(\zeta)$ is positive definite and a soliton exists, the question of its stability is far from trivial. In this paper we establish that a soliton is stable if $L(\zeta)$ is a positive definite fourth-order polynomial. The proof of its stability is based on the boundedness of the Hamiltonian for a fixed adiabatic invariant. We assume that the same conclusion regarding the stability will be valid for any positive definite polynomial $L(\zeta)>0$ of even degree. However, if we have

$$|F(\zeta)|<C|\zeta|^{\alpha} \quad \text{for } |\zeta|\to\infty \tag{1.4}$$

and $\alpha\leqslant 1/2$, stability of the soliton is doubtful, and it is more likely unstable.

There is one more important point on which we would like to focus attention in this article. The objects which have traditionally been called solitons in nonlinear optics are not such in the strict sense of the word. They are quasisolitons, i.e., approximate solutions of Maxwell's equations, which depend on four parameters. Real stationary solitons, which propagate with a constant velocity without changing their form, are exact solutions of Maxwell's equations, which depend on two parameters. The latter exist, if the dielectric constant $\varepsilon(\omega)$ has a maximum in the frequency range under consideration for a focusing nonlinearity or a minimum, if the medium is defocusing. In a purely conservative medium quasisolitons exist for a finite time owing to radiation as a result of multiphoton processes. In reality, however, this time is much greater than the lifetime resulting from the linear damping, and the difference between solitons and quasisolitons is insignificant.

2. STATIONARY SOLITONS

In this section we demonstrate how to find a soliton solution directly from Maxwell's equations. We consider a very simple model of the simultaneous propagation of pulses, assuming that the polarization is linear and that the electric field $E(x,t)$ is perpendicular to the propagation direction. In this case Maxwell's equations can be reduced to the wave equation for the field $E(x,t)$:

$$\frac{\partial^2 D}{\partial t^2}-c^2\frac{\partial^2 E}{\partial x^2}=0, \tag{2.1}$$

where the electric displacement D is assumed to be related to the electric field by the expression

$$D(x,t)=\hat{\varepsilon}(t)E(x,t)+\chi E^3(x,t). \tag{2.2}$$

In this expression $\hat{\varepsilon}$ is an integral operator; the Fourier transform of its kernel is $\varepsilon(\omega)$, i.e., the dielectric constant. The second term in (2.2) corresponds to the Kerr effect, and χ is the Kerr constant.

The function $\varepsilon(\omega)$ is analytically continuable into the upper half-plane of ω (see, for example, Ref. 7). For real values of ω the magnitude of $\varepsilon(\omega)$ obeys the Kramers–Kronig relations. It particular, it follows from these relations that on the real axis the imaginary part of the dielectric constant ε'', which is responsible for the dissipation of electromagnetic waves, cannot be equal to zero at all frequencies. Below we shall assume that there is a certain frequency band $\Delta\omega$, within which the imaginary part of the dielectric constant is small enough that it can be neglected.

Let us consider the propagation of a wave packet with a spectrum lying within this transparency window, assuming that the frequency width of the pulse spectrum is small compared with $\Delta\omega$. A solution in the form of an isolated pulse, i.e., a soliton, can be expected only under such conditions.

As was noted in the Introduction, two types of solitons are possible. The solitons of the first type are stationary in a moving frame. They propagate with a constant velocity without changing their form. A classical example of solitons of this type is provided by the solitons for the KdV equation, which, in particular, describe solitary waves in shallow water. The solitons of the other type are called quasisolitons. They have internal dynamics and propagate with a constant velocity only on the average. The classical quasisolitons include breezers, which are described by the sine-Gordon equation (for further information, see, for example, Refs. 8–10).

Stationary solitons are exact solutions of Eq. (2.1). We shall seek these solutions in the form

$$E=E(x-vt), \tag{2.3}$$

where v is the constant velocity and E tends to zero at infinity. The substitution of (2.3) into (2.1) makes it possible to integrate the equation twice:

$$\hat{L}E(x)=\alpha E^3(x), \quad \alpha=\chi v^2/c^2, \tag{2.4}$$

where the operator $\hat{L}$ equals

$$\hat{L}=1-\frac{v^2}{c^2}\,\hat{\varepsilon}. \tag{2.5}$$

In the Fourier representation $\hat{L}$ is written in the form

$$L(\omega)=1-\frac{v^2\varepsilon(\omega)}{c^2}, \tag{2.6}$$

where the frequency ω and the wave number k are related by the equality $\omega=kv$. The second term in (2.6) is the square of the ratio between v and the phase velocity of an electromagnetic wave of small amplitude:

$$v_{\rm ph}=c/\sqrt{\varepsilon(\omega)}. \tag{2.7}$$

Hence it is easily seen that the operator $\hat{L}$ becomes positive definite if and only if

$$v_{\rm ph}^2(\omega)>v^2, \tag{2.8}$$

for all ω, and it accordingly becomes negative definite in the opposite case:

$$v_{\rm ph}^2(\omega)<v^2. \tag{2.9}$$

We now show that a soliton solution is possible only when condition (2.8) or (2.9) is satisfied. Let us assume that the opposite is true, i.e., let the conditions (2.8) and (2.9) not be satisfied. In this case the equation

$$\frac{v^2\varepsilon(\omega)}{c^2}=1 \tag{2.10}$$

has a solution (for simplicity we assume that it is unique: $\omega=\omega_0$). Then Eq. (2.3) can be rewritten in the following manner:

$$E(x-vt)=E_0(x-vt)+\hat{L}^{-1}(1-\hat{P})\alpha E^3(x-vt). \tag{2.11}$$

Here

$$E_0(x-vt)=\mathrm{Re}(A\ \exp[-i\omega_0(t-x/v)])$$

is the solution of the homogeneous linear equation

$$\hat{L}E_0=0, \tag{2.12}$$

and $\hat{P}$ is a projector onto the state $E_0(x-vt)$, so that $(1-\hat{P})\chi E^3(x-vt)$ is orthogonal to E_0 and, therefore, the operator $\hat{L}$ is reversible in this class of functions. To find the explicit solution of Eq. (2.11), we can use, for example, an iterative scheme, taking E_0 as the zeroth approximation. It is of fundamental significance that, by proceeding in this manner, we must arrive at nonlocalized solutions, which depend on two parameters, viz., the imaginary and real parts of the complex amplitude A. Hence the following conclusion can be drawn: the stationary equation (2.3) can have a soliton solution if $\hat{L}$ is sign-definite. If Eq. (2.12) has a nontrivial solution, or, equivalently, if the phase velocity v_{ph} and the velocity v are equal, i.e., if

$$v_{\mathrm{ph}}=v, \tag{2.13}$$

there is no stationary soliton solution. We note that this conclusion relies heavily on the fact that the singularity on the right-hand side of Eq. (2.11) $(E^3)_\omega/L(\omega)$ is not removable. As will be shown below, singularities of this type can be removed, if the matrix element of the four-wave interaction (χ in the present case) has a frequency dependence.

Equation (2.13) can also be regarded as a condition for Cherenkov radiation by a moving object. The nature of the object itself is not important here. It can be a charged particle, a ship, or, for example, a soliton. In any case the moving object loses energy as a result of Cherenkov radiation. In the case under discussion this means that if the velocity of an electromagnetic soliton satisfies the conditions (2.9), it must emit waves, and, therefore, such a pulse cannot exist as a stationary object. Thus, we arrive at the following condition for the existence of solitons: a soliton solution can exist when the equation

$$\omega(k)=kv \tag{2.14}$$

does not have a (real) solution. Here $\omega=\omega(k)$ is the dispersion law. For electromagnetic waves $\omega(k)$ is determined from the equation

$$\omega^2=k^2c^2/\varepsilon(\omega). \tag{2.15}$$

The relation (2.14) has a simple interpretation in the $\omega-k$ plane. The right-hand side of (2.14) corresponds to a straight line emerging from the origin of coordinates, and, accordingly, the velocity v in this plane equals the slope $\tan\phi$:

$$v=\tan\phi.$$

The existence of a solution for Eq. (2.14) is indicated by the intersection of the $\omega=\omega(k)$ curve by the straight lines. This assigns a complete cone of angles Ω, where stationary soliton solutions are impossible. Cone $\tilde{\Omega}$, which is complementary to Ω, corresponds to possible soliton solutions. On the boundary $\partial\Omega$ between the cones the straight lines are tangent to the $\omega=\omega(k)$ dispersion curve, and at the points of tangency k_i the group and phase velocities coincide:

$$\left.\frac{\omega(k)}{k}\right|_{k_i}=\left.\frac{\partial\omega(k)}{\partial k}\right|_{k_i}. \tag{2.16}$$

For the dispersion law (2.15) this relation is written as

$$\left.\frac{d\varepsilon(\omega)}{d\omega}\right|_{\omega_i}=0. \tag{2.17}$$

It is natural to assume that the soliton amplitude vanishes at these critical points (since there should not be any stationary soliton solutions outside $\tilde{\Omega}$). As will be shown below, the behavior of a soliton solution near these critical points is universal. We demonstrate this fact in the case of the stationary equation (2.3). It is, however, fundamentally important that the result is general and can be used for other models. This fact was first investigated for capillary-gravitational solitons in deep water.[11–13] The spectrum of capillary-gravitational waves is known to have a minimum phase velocity for wave numbers lying in the intermediate region between the gravitational and capillary portions of the spectrum.

For simplicity, we assume that Eq. (2.17) has only one positive solution $\omega=\omega_0$ [because of the parity of $\varepsilon(\omega)$ there is one more root $\omega=-\omega_0$], and let the cone of angles $\tilde{\Omega}$ lie below the critical velocity:

$$v<v_{\mathrm{cr}}=\frac{c}{\sqrt{\varepsilon(\omega_0)}}.$$

Thus, the function $\varepsilon(\omega)$ has two identical maxima at symmetric points, and

$$\frac{d^2\varepsilon(\pm\omega_0)}{d\omega^2}<0.$$

In this case $\hat{L}$ is an invertible operator, and Eq. (2.4) can be written in the form

$$E_\omega=\frac{1}{L(\omega)}\alpha(E^3)_\omega. \tag{2.18}$$

Near the critical velocity ($v_{\mathrm{cr}}-v\ll v_{\mathrm{cr}}$) the plot of $L(\omega)$ as a function of ω is close to zero in small vicinities of the two points $\omega=\pm\omega_0$ because of its symmetry with respect to ω. Therefore, according to (2.18) the distribution of $E(\omega)$ is determined to a considerable extent by the function $1/L(\omega)$. Accordingly, in the t-representation the solution will be close to a monochromatic wave. It is important that the monochromaticity of the wave improves as v approaches v_{cr}. Therefore, $E(t')$ $(t'=t-x/v)$ will be sought in the form of an expansion in the harmonics $n\omega_0$:

$$E(t)=\sum_{n=0}^{\infty}[E_{2n+1}(\tau)e^{-i(2n+1)\omega_0 t'}+\text{c.c.}]. \tag{2.19}$$

Here we have formally introduced the small parameter

$$\epsilon=\sqrt{1-v/v_{\text{cr}}} \tag{2.20}$$

and the slow time $\tau=\epsilon t'$, so that the $E_{2n+1}(\tau)$ are the envelope amplitudes of each harmonic. The representation (2.19) means that the width of each harmonic along the frequency scale, $\delta\omega\sim\epsilon$, is small compared with the frequency ω_0, i.e., the Fourier spectrum (2.19) is a series of narrow peaks. The main peaks correspond to the first harmonic. Therefore, the action of $\hat{L}$ on (2.19) can be expanded into a series in powers of ϵ. Assuming that $E_{2n+1}\sim\epsilon^{2n+1}$ and substituting (2.19) into the stationary equation (2.4), with consideration of (2.17) in the first order we arrive at a stationary nonlinear Schrödinger equation:

$$\epsilon^2 E_1-S\frac{\partial^2 E_1}{\partial t^2}-\frac{3}{2}\alpha|E_1|^2E_1=0, \tag{2.21}$$

where

$$S=-\frac{v^2}{4c^2}\frac{d^2\epsilon(\omega_0)}{d\omega^2}>0. \tag{2.22}$$

Equation (2.21) has a soliton solution only if $\alpha>0$:

$$E_1(t')=\frac{2\epsilon}{\sqrt{3\alpha}}\,\text{sech}\left[\frac{\epsilon(t-x/v-t_0)}{\sqrt{S}}\right]. \tag{2.23}$$

This solution is unique to within a constant phase multiplier. It is the universal asymptote of the soliton solution. As v approaches v_{cr}, its amplitude vanishes according to a square-root law $\sim\sqrt{v_{\text{cr}}-v}$, and the soliton pulse width Δt increases in inverse proportion to this factor:

$$\Delta t=\sqrt{S}/\epsilon.$$

For times greater than Δt we must take into account the following expansion terms, particularly the third-order dispersion and the corrections to the cubic nonlinearity. In this time range the soliton behavior is no longer universal.

It is noteworthy that Eq. (2.21) does not have solitonlike solutions when $\epsilon^2=1-v/v_{\text{cr}}<0$.

When the tangent approaches the dispersion curve from above, S becomes negative. For this reason solitons exist only for defocusing media ($\alpha<0$).

The case where the point of tangency satisfies $\omega_0=0$ calls for a special treatment. Near the critical velocity the stationary equation (2.3) does not require the expansion (2.19). It is sufficient to expand $\epsilon(\omega)$ near $\omega=0$:

$$\varepsilon(\omega)=\varepsilon(0)+\frac{1}{2}\frac{d^2\varepsilon(0)}{d\omega^2}\omega^2.$$

According to this expansion, the stationary equation takes on the form

$$\epsilon^2 E-S\partial_t^2 E-\frac{1}{2}\alpha E^3=0, \tag{2.24}$$

where, as before, ϵ, S, and α are given, respectively, by Eqs. (2.20), (2.22), and (2.4) taken at $\omega=0$. The localized solution of Eq. (2.24) has the form of a soliton for the modified Korteweg-de Vries (MKdV) equation:

$$E(t-x/v)=\frac{2\epsilon}{\sqrt{\alpha}}\,\text{sech}\left[\frac{\epsilon(t-x/v-t_0)}{\sqrt{S}}\right].$$

3. QUASISOLITONS; HIGHER-ORDER DISPERSION

In this section we discuss the difference between solitons and quasisolitons in the case of a generalized nonlinear Schrödinger equation (GNLSE). The GNLSE has a more extensive class of soliton solutions than does the original Maxwell equation. Unlike the stationary solitons (2.23), these solutions are approximate and depend on four parameters. However, the mechanism for selecting the soliton solutions remains the same as for the stationary solitons considered in the preceding section.

The transparency window $\Delta\omega$ must be small compared with the mean value of the frequency ω_0: $\omega_0\gg\Delta\omega$. In this case an envelope can be introduced for the entire region. The most convenient and systematic approach for obtaining the equation for the envelopes is based on the Hamiltonian formalism.[14]

Let us consider Eq. (2.1), which we present in the form of a system of equations:

$$\frac{\partial\rho}{\partial x}+\frac{\partial^2\phi}{\partial t^2}=0,\quad \frac{\partial\phi}{\partial x}+\frac{1}{c^2}\left(\hat{\varepsilon}\rho+\frac{4\pi\chi}{c^2}\rho^3\right)=0. \tag{3.1}$$

The potential ϕ and the ''density'' ρ introduced here are related to the electric field E and the magnetic field H by the expressions

$$E=\frac{\sqrt{4\pi}}{c}\rho,\quad H=\sqrt{4\pi}\,\frac{\partial\phi}{\partial t}. \tag{3.2}$$

Equations (3.1) can be written in Hamiltonian form:

$$\frac{\partial\rho}{\partial x}=\frac{\delta\mathscr{H}}{\delta\phi},\quad \frac{\partial\phi}{\partial x}=-\frac{\delta\mathscr{H}}{\delta\rho}. \tag{3.3}$$

Here x plays the role of the time, and the Hamiltonian has the form of an integral with respect to time:

$$\mathscr{H}=\int\left[\frac{1}{2}\left(\frac{\partial\phi}{\partial t}\right)^2+\frac{1}{2c^2}\rho\hat{\varepsilon}\rho+\frac{\pi\chi}{c^4}\rho^4\right]dt$$

$$\equiv\frac{1}{8\pi}\int\left[H^2+E\hat{\varepsilon}E+\frac{1}{2}\chi E^4\right]dt. \tag{3.4}$$

The quadratic part of $\mathscr{H}$ defines a linear dispersion law for $k=k(\omega)$, which coincides with (2.15). We can go over to the normal variables $a_\omega(x)$ using the replacements

$$\rho_\omega=\sqrt{\frac{\omega^2}{2k(\omega)}}\,(a_\omega^*+a_{-\omega}),$$

$$\phi_\omega=-i\sqrt{\frac{k(\omega)}{2\omega^2}}\,(a_\omega^*-a_{-\omega}), \tag{3.5}$$

where ρ_ω and ϕ_ω are the Fourier transforms of the density ρ and the potential ϕ, and $k(\omega)$ is understood in these formu-

las as a positive root of the dispersion relation (2.15). The substitution of these relations into Eq. (3.3) gives the equations of motion in the variables a_ω:

$$\frac{\partial a_\omega}{\partial x} = i\frac{\delta \mathscr{H}}{\delta a_\omega^*}, \tag{3.6}$$

where the Hamiltonian $\mathscr{H}$ takes the standard form (compare Ref. 14):

$$\mathscr{H} = \int k(\omega)|a_\omega|^2 d\omega + \frac{1}{2}\int T_{\omega_1\omega_2\omega_3\omega_4} a_{\omega_1}^* a_{\omega_2}^* a_{\omega_3} a_{\omega_4} \times \delta_{\omega_1+\omega_2-\omega_3-\omega_4}\Pi_i d\omega_i . \tag{3.7}$$

The matrix element T appearing therein is assigned by the formula

$$T_{\omega_1\omega_2\omega_3\omega_4} = \frac{3\chi}{4\pi c^4}\left[\frac{\omega_1^2\omega_2^2\omega_3^2\omega_4^2}{k(\omega_1)k(\omega_2)k(\omega_3)k(\omega_4)}\right]^{1/2}. \tag{3.8}$$

If the fourth-order susceptibility χ depends on the frequencies, the constant χ in the matrix element (3.8) is replaced by $\chi(\omega_1\omega_2\omega_3\omega_4)$ with the necessary symmetry properties (see Refs. 7 and 15), which ensure the following symmetry relations for T:

$$T_{\omega_1\omega_2\omega_3\omega_4} = T_{\omega_2\omega_1\omega_3\omega_4} = T_{\omega_1\omega_2\omega_4\omega_3} = T^*_{\omega_3\omega_4\omega_1\omega_2}. \tag{3.9}$$

In the Hamiltonian (3.7) we retained only the terms responsible for the scattering of waves, neglecting all the other processes, which make a contribution in the next (sixth) order with respect to the amplitude of the waves for narrow wave packets.

The Hamiltonian formulation of the equations of motion (3.6) guarantees ''conservation'' (absence of a dependence on x) of the Hamiltonian $\mathscr{H}$, as well as of the ''momentum''

$$P = \int \omega|a_\omega|^2 d\omega, \tag{3.10}$$

which coincides exactly with the Poynting vector integrated over time:

$$P = \frac{c}{4\pi}\int_{-\infty}^{\infty} EH dt.$$

Let us now proceed to the derivation of the equation for the envelopes by introducing the packet envelope amplitude:

$$\psi(t,x) = \frac{1}{\sqrt{2\pi}}\int a_\omega e^{-i(\omega-\omega_0)t - ik_0(\omega_0)x} d\omega.$$

Here we assume that the spectrum of a_ω is concentrated in a narrow interval $\delta\omega$ near ω_0 and that $\omega_0 \gg \delta\omega$. Accordingly, $\psi(t,x)$ is a slow function of the coordinates and the time.

Next, expanding $k(\omega)$ and $T_{\omega_1\omega_2\omega_3\omega_4}$ into a series in $\Omega = \omega - \omega_0$ at ω_0 we have

$$\kappa(\Omega) = k(\omega) - k(\omega_0) = \frac{1}{v_{\rm gr}}\Omega - k_0 S\Omega^2 - \gamma\Omega^3 + \delta\Omega^4 + \ldots, \tag{3.11}$$

$$T_{\omega_1\omega_2\omega_3\omega_4} = T_0 + \frac{\partial T}{\partial\omega_1}(\Omega_1+\Omega_2+\Omega_3+\Omega_4) + \frac{1}{2}\frac{\partial^2 T}{\partial\omega_1^2}(\Omega_1^2+\Omega_2^2+\Omega_3^2+\Omega_4^2) + \frac{\partial^2 T}{\partial\omega_1\partial\omega_2}(\Omega_1\Omega_2+\Omega_3\Omega_4) + \frac{\partial T}{\partial\omega_1\partial\omega_3}(\Omega_1\Omega_3+\Omega_1\Omega_4+\Omega_2\Omega_3+\Omega_2\Omega_4) + \ldots . \tag{3.12}$$

In the expression for $k(\omega)$ we have retained the terms up to fourth order in Ω, and in the matrix element T we have retained the terms that are quadratic in Ω. In expanding the matrix element, for simplicity, we considered it to be real and utilized its symmetry properties (3.9). Accordingly, the coefficients in (3.12) are

$$T_0 = T_{\omega_0\omega_0\omega_0\omega_0}, \quad \frac{\partial T}{\partial\omega_1} = \left.\frac{\partial T_{\omega_1\omega_2\omega_3\omega_4}}{\partial\omega_1}\right|_{\omega_k=\omega_0},$$

$$\frac{\partial^2 T}{\partial\omega_i\partial\omega_j} = \left.\frac{\partial^2 T_{\omega_1\omega_2\omega_3\omega_4}}{\partial\omega_i\partial\omega_j}\right|_{\omega_k=\omega_0}.$$

Next, performing the inverse Fourier transformation with respect to Ω, for ψ we obtain the generalized nonlinear Schrödinger equation

$$i\left(\frac{\partial\psi}{\partial x} + \frac{1}{v_{\rm gr}}\frac{\partial\psi}{\partial t}\right) + K_0 S\psi_{tt} + \beta_1|\psi|^2\psi = -i\gamma\psi_{ttt} - 4i\beta_2|\psi|^2\psi_t - \delta\psi_{tttt} + (\beta_3-\beta_4)[(\psi^2\psi_t^*)_t - (\psi_t)^2\psi^*] + (\beta_3+\beta_5)\psi^*(\psi^2)_{tt} - \beta_6|\psi|^4\psi. \tag{3.13}$$

The left-hand side of this equation corresponds to the classical nonlinear Schrödinger equation: the second term in it describes the propagation of a wave packet as a whole and, therefore, can be eliminated by going over to the local coordinate frame. The next term ($\sim S$) is responsible for quadratic dispersion. Now, for $d\epsilon(\omega_0)/d\omega_0 = 0$ the coefficient S coincides with the expression in (2.22). The last term on the left-hand side defines a nonlinear correction to the frequency of the monochromatic wave. The first two terms on the right-hand side are $\sim(\delta\omega/\omega_0)^3$. It is important that there are only two such terms. In this case the coefficient $\beta_2 = 2\pi\partial T/\partial\omega$ is nonzero even for a constant fourth-order susceptibility χ. When $\chi = {\rm const}$ holds, β_2 can vanish only if $k\sim\omega^2$. The remaining terms are $\sim(\delta\omega/\omega)^4$. Among them we took into account the terms $\sim|\psi|^4\psi$, which are of the same order of magnitude.

The coefficients β_i appearing in Eq. (3.13) take on a very simple form for the matrix element (3.8):

$$\beta_1 = \frac{3}{2}k_0^2\chi\left(\frac{v_{\rm ph}}{c}\right)^4, \quad \beta_2 = \frac{\beta_1}{\omega_0}\left(1 - \frac{v_{\rm ph}}{2v_{\rm gr}}\right),$$

$$\beta_3=\beta_1\frac{k^{1/2}}{\omega_0}\frac{\partial^2}{\partial\omega_0^2}\left(\frac{\omega_0}{k^{1/2}}\right),\quad \beta_4=\beta_5=\frac{\beta_1}{\omega_0^2}\left(1-\frac{v_{\rm ph}}{2v_{\rm gr}}\right)^2. \tag{3.14}$$

According to its derivation, Eq. (3.13) should be classified as a Hamiltonian equation:

$$i\frac{\partial\psi}{\partial x}=-\frac{\delta\mathscr{H}}{\delta\psi^*}. \tag{3.15}$$

Here the Hamiltonian $\mathscr{H}$ can be represented in the form of a sum of Hamiltonians:

$$\mathscr{H}=\mathscr{H}_1+\mathscr{H}_2+\mathscr{H}_3+\mathscr{H}_4+\ldots,$$

where

$$\mathscr{H}_1=\frac{i}{v_{\rm gr}}\int\psi^*\psi_t dt, \tag{3.16}$$

$$\mathscr{H}_2=-\int\left(k_0S|\psi_t|^2-\frac{\beta_1}{2}|\psi|^4\right)dt, \tag{3.17}$$

$$\mathscr{H}_3=\int\{i\gamma\psi^*\psi_{ttt}+i\beta_2(\psi^*\psi_t-\psi\psi_t^*)|\psi|^2\}dt, \tag{3.18}$$

$$\mathscr{H}_4=\int\left\{\delta|\psi_{tt}|^2-\frac{\beta_3}{2}|\psi|^2(\psi\psi_{tt}^*+\text{c.c.})-\frac{\beta_4}{2}(\psi_t^2\psi^{*2}+\text{c.c.})-\frac{\beta_5}{2}\psi^{*2}\partial_t^2\psi^2+\frac{\beta_6}{3}|\psi|^6\right\}dt. \tag{3.19}$$

Here $\mathscr{H}_2$ corresponds to the classical NLSE, and the next Hamiltonian corresponds to the complex MKdV equation. It is important that each of the successive Hamiltonians is smaller than the preceding one. However, this situation can change, if any of the expansion coefficients introduces additional smallness. As is seen from (2.23), the soliton width decreases as the quadratic dispersion coefficient S decreases. Therefore, when S is small (such a situation arises near the so-called zero-dispersion point), the cubic dispersion ($\sim\gamma$) must be taken into account with neglect of all the higher-order terms, as well as the term that is proportional to β_2. If β_1 is small, the nonlinear dispersion, which is proportional to β_2, must be taken into account with neglect of the cubic linear dispersion.

Let us now turn to an analysis of the solitonlike solutions for the generalized Schrödinger equation.

To illustrate how the mechanism (2.23) operates, we first consider the nonlinear Schrödinger equation with quadratic dispersion [which corresponds to the Hamiltonian (3.17)]:

$$i\frac{\partial E}{\partial x}+E_{tt}+2|E|^2E=0. \tag{3.20}$$

Here we have used dimensionless variables, and the nonlinearity is assumed to be focusing, $S\alpha>0$.

It is noteworthy that, unlike the wave equation (2.1), a generalized NLSE, particularly the NLSE with quadratic dispersion, has an additional symmetry, viz., $E\to Ee^{i\phi}$, which appears as a result of the averaging of the equations of fast oscillations. Therefore, the envelope soliton solutions form a more extensive class of solutions than does the wave equation (2.1). According to our definition, they should be classified as quasisolitons. To find the corresponding solutions, we should set $E(x,t)=e^{i\lambda x}\psi(t+\beta x)$, where ψ obeys the equation[1)]

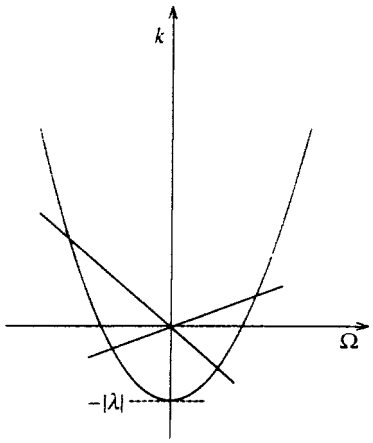

FIG. 1. Dispersion curve (3.23) for negative λ. Any straight line emerging from the origin of coordinates intersects the dispersion curve.

$$L(i\partial_t)\psi\equiv-i\beta\psi_t+\lambda\psi-\psi_{tt}=2|\psi|^2\psi. \tag{3.21}$$

In the case under consideration the conditions for Cherenkov radiation (2.14) are written in the following manner:

$$\beta\Omega=k(\Omega)\ \text{ or }\ L(\Omega)=0, \tag{3.22}$$

where the dispersion relation for Eq. (3.21) takes the form

$$k(\Omega)=\lambda+\Omega^2. \tag{3.23}$$

Hence it is seen that for $\lambda<0$ the resonance condition (3.22) is satisfied for any value of the velocity (Fig. 1), and hence no solitons exist in this case. This is verified directly by solving Eq. (3.21): for $\lambda<0$ all the solutions are periodic or quasiperiodic. Soliton solutions are possible only for positive values of λ. Their velocities lie in the range $-2\sqrt{\lambda}\leqslant\beta\leqslant2\sqrt{\lambda}$ (Fig. 2). At the points $\Omega=\pm\sqrt{\lambda}$ the straight line $k=\beta_{\rm cr}\Omega$ is tangent to the $k=k(\Omega)$ dispersion curve. According to the results of Sec. 2, the soliton solution should vanish at these points, as follows directly from the solution of Eq. (3.21):

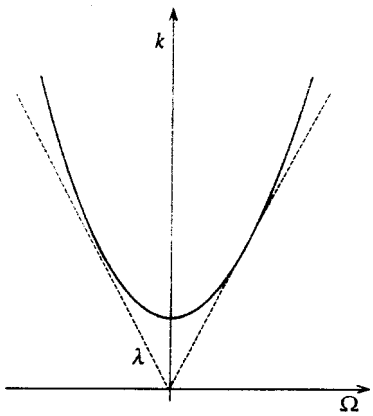

FIG. 2. Dispersion curve (3.23) for positive λ. The dashed lines which are tangent to the dispersion curve correspond to the critical velocities $\beta=\pm2\sqrt{\lambda}$. These straight lines specify the boundary of the soliton cone of angles.

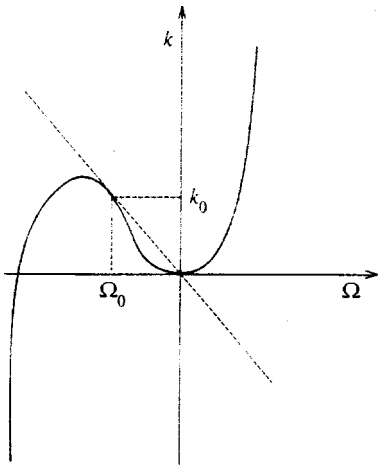

FIG. 3. Third-order dispersion $k=S\Omega^2+\gamma\Omega^3$. The dashed straight line is tangent to the dispersion curve at $\Omega=\Omega_0$, but intersects it at $\Omega=0$.

$$E=e^{i\lambda x}\frac{e^{i\beta t'}\Delta\Omega}{\cosh(\Delta\Omega t')},\quad \Delta\Omega=\sqrt{\lambda-\frac{\beta^2}{4}}. \tag{3.24}$$

Hence, the region for the existence of solitons is given by the inequality $\lambda>\beta^2/4$. The upper bound in this inequality specifies the critical velocity

$$\beta_{cr}=\pm 2\sqrt{\lambda}.$$

It is important to note that the operator L in Eq. (3.21) is positive definite for $\lambda>\beta^2/4$.

Let us now turn to the third-order dispersion. We assume, as before, that the soliton solution contains an exponential multiplier

$$E(x,t)=e^{i\lambda x}\psi(t'),\quad t'=t+\beta x. \tag{3.25}$$

The corresponding operator $L(i\partial_t)$ has the form

$$L(\Omega)=-\beta\Omega+\lambda+S\Omega^2+\gamma\Omega^3. \tag{3.26}$$

This operator is sign-definite for any values of λ, β, S, and $\gamma\neq 0$. This means that the equation $L(\Omega)=0$ or the equivalent equation

$$\beta\Omega=\lambda+S\Omega^2+\gamma\Omega^3,$$

has at least one real solution: the dispersion curve for $k(\Omega)=\lambda+S\Omega^2+\gamma\Omega^3$ always intersects any straight line emerging from the origin of coordinates. For example, for $\lambda=0$ and $\beta\geqslant\beta_0=-S^2/(4\gamma)$ all the straight lines $k=\beta\Omega$ intersect the $k=k(\Omega)$ dispersion curve twice. For $\beta<\beta_0$ the straight lines have one point of intersection, and for $\beta=\beta_0$ tangency occurs (Fig. 3). However, one point of intersection is sufficient for the absence of solitons. On the other hand, the example of the KdV equation, which simultaneously has cubic dispersion and solitons, apparently contradicts the foregoing statement. Actually, there is no contradiction here. Everything is explained by the dependence of the matrix element on the wave vector, which provides for cancellation of the singularity in the equation of the form (2.18).

We can show in the example of the KdV equation

$$U_t+U_{xxx}+6UU_x=0, \tag{3.27}$$

how cancellation of a singularity occurs. For a soliton moving with the velocity v,

$$L(k)=ik(v+k^2).$$

For $v>0$ the equation $L(k)=0$ has one real root $k=0$. In this case the analog of Eq. (2.18) is

$$U_k=\frac{3ik(U^2)_k}{L(k)},$$

which clearly does not contain a singularity at $k=0$. The situation is similar for other equations of the KdV type (see, for example, Ref. 16).

Solutions of the soliton type were recently obtained[17] for a generalized Schrödinger equation, which simultaneously takes into account the third-order dispersion and corresponds to its nonlinearity [in the present paper this corresponds to consideration of the Hamiltonians (3.17) and (3.18)]. If the relations between γ and β_2 are arbitrary, the soliton solution found in Ref. 17 has a spectrum concentrated at the frequencies $\Omega\sim 1/\gamma$, $1/\beta_2$, i.e., at frequencies comparable to ω_0. In the unique case where the relation between the coefficients has the form

$$\frac{K_0 S}{\beta_1}=\frac{3\gamma}{4\beta_2},$$

the soliton spectrum is displaced by a small amount. This case is special, i.e., Eq. (3.13) (written in dimensionless variables),

$$iE_x+E_{tt}+2|E|^2E=i\epsilon(E_{ttt}+6|E|^2E_t), \tag{3.28}$$

allows application of the inverse scattering problem technique (see, for example, Ref. 8). In this case the Hamiltonians (3.17) and (3.18) are conserved independently. They are both created by the same associated operator, viz., the Zakharov-Shabat operator.[2] The parameter ϵ in this equation is of order $\delta\omega/\omega$, and E takes values of order unity. Soliton solutions for this equation were first pointed out in Ref. 18. The simplest of them is the solution

$$E=e^{i\mu^2 x}\frac{\mu}{Ch\,\mu(t-\epsilon\mu^2 x)},$$

which transforms into a stationary soliton of the NLSE (3.24) when $\epsilon=0$.

One conclusion which can be drawn from the foregoing material is that the existence of soliton solutions for the third-order operators L is due to the presence of derivatives in the nonlinear term or, stated differently, the dependence of the matrix elements on the frequency. If there is no such dependence, or if it is insignificant, as is the case, for example, near the point of zero dispersion, there are no reasons for cancellation of the singularities in the equation of the form (2.18). Therefore, the results in Ref. 19 of the numerical observation of solitons for the NLSE with cubic dispersion should be revised (see also Ref. 20, which was devoted to this equation).

We shall henceforth confine ourselves to consideration of the case where there is no dispersion of the nonlinearity or

it is insignificant. In such a situation third-order dispersion cannot provide for the existence of solitons, i.e., the next expansion terms must be taken into account.

For fourth-order dispersion the corresponding operator L has the form

$$L(\Omega) = -\beta\Omega + \lambda + S\Omega^2 + \gamma\Omega^3 + \delta\Omega^4. \tag{3.29}$$

The sign-definiteness of L is now determined by the sign of δ: the operator is positive definite for $\delta > 0$ and negative definite in the opposite case.

The cubic term can always be eliminated from L by means of an appropriate frequency shift $\Omega \to \Omega + \nu$. Furthermore, using simple scaling and sign reversal, $L(\Omega)$ can be brought into the following two canonical forms:

$$L(\Omega) = -\beta\Omega + k(\Omega) = -\beta\Omega + \lambda + (\Omega^2 - \gamma_0^2)^2, \tag{3.30}$$

$$L(\Omega) = -\beta\Omega + k(\Omega) = -\beta\Omega + \lambda + (\Omega^2 + \nu_0^2)^2. \tag{3.31}$$

Then, applying the criterion (3.22) to the dispersion law (3.30) with $\lambda < 0$, we can easily see that the resonance condition (3.22) is satisfied for all values of β and that the existence of solitons is, therefore, impossible in this region of parameters.

For positive $\lambda = \mu^4$ solitons are possible in the region $-\beta_{cr} \leqslant \beta \leqslant \beta_{cr}$, where

$$\beta_{cr} = 4\Omega_0(\Omega_0^2 - \nu_0^2) \quad \text{and} \quad \Omega_0^2 = \frac{1}{6}(2\nu_0^2 + \sqrt{16\nu_0^4 + 12\mu^4}). \tag{3.32}$$

Near the critical velocity (3.32) the dispersion is positive; therefore, localized solutions of the soliton type can exist only for focusing ($\delta\chi > 0$) nonlinearity, while nonlinearity with respect to the quadratic dispersion would be defocusing. The form of the soliton in this case is determined from the equation

$$L(i\partial_t)\psi = 2\sigma|\psi|^2\psi, \tag{3.33}$$

where $L(i\partial_t)$ is given by Eq. (3.30) or (3.31), $\sigma = \mathrm{sgn}(\delta\chi)$ specifies the character of the nonlinear interaction: for $\sigma = 1$ it is attractive, and for $\sigma = -1$ it is repulsive. Soliton solutions are possible only for a focusing medium.

The simplest solutions of (3.33) are stationary solitons. Their form is found by integrating the equation

$$\mu^4\psi + (\partial_t^2 + \nu_0^2)^2\psi - 2|\psi|^2\psi = 0. \tag{3.34}$$

It is significant that a moving soliton for fourth-order dispersion has a profile which differs from a soliton for the NLSE with quadratic dispersion. It cannot be deformed into a stationary soliton by simple scaling and phase transformation.

To find the solution, Eq. (3.34) must be supplemented by the boundary conditions

$$\psi, \ \psi_t \to 0 \ \text{ as } \ t \to \pm\infty.$$

The symmetry of Eq. (3.34) allows real symmetric (relative to t) solutions: $\psi(t) = \psi(-t) = \psi^*(t)$. At infinity ($t \to \pm\infty$) these solutions should decay exponentially: $\psi \sim e^{\nu t} \to 0$, where the exponent ν is determined from the equation

$$\nu^4 + (\nu^2 + \nu_0^2)^2 + \mu^4 = 0.$$

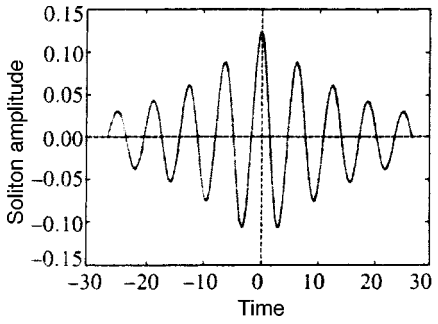

FIG. 4. Dependence of the soliton amplitude (in units of ν_0^2) on the time (in units of ν_0^{-1}) for $\mu/\nu_0 = 1/3$. The soliton envelope has the form of the function sech to good accuracy.

The roots of this equation are assigned by the expressions

$$\nu = \pm\left[\frac{1}{2}(\sqrt{\mu^4 + \nu_0^4} - \nu_0^2)\right]^{1/2} \pm i\left[\frac{1}{2}(\sqrt{\mu^4 + \nu_0^4} - \nu_0^2)\right]^{1/2}. \tag{3.35}$$

They are all complex. This means, in particular, that all stationary solitons should have an oscillating structure. If $\mu \sim \nu_0$ holds, the real and imaginary parts of the exponent ν are of the same order. Critical tangency occurs when $\mu = 0$. Near this point the real part of ν' is small for a finite value of the imaginary part:

$$\nu = \pm\mu^2/\nu_0 \pm i\nu_0. \tag{3.36}$$

Envelope solitons of the universal form (2.23) appear in just this limit.

For large μ ($\mu \gg \nu_0$) the roots have the asymptote

$$\nu = \mu\frac{\pm 1 \pm i}{\sqrt{2}}.$$

Figures 4–6 show the solitons for different values of μ and ν_0. In the limit $\mu \to 0$ (Fig. 4) the soliton has a clearly expressed envelope soliton form, and at large μ ($\mu \gg \nu_0$) the soliton has only one oscillation on its scale (Fig. 6). At large distances (large times) the solitons for all the values of μ and ν_0 have exponentially decaying, oscillating tails. As the ratio μ/ν_0 increases, the amplitude of the soliton increases, and its width decreases. The solitons obtained here, like the real

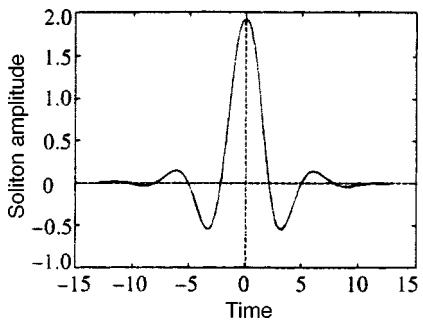

FIG. 5. Form of a soliton when $\mu/\nu_0 = 1$. The amplitude of the soliton (in units of ν_0^2) increases, and its width (in units of ν_0^{-1}) decreases. Oscillations are still observed on the scale of the soliton.

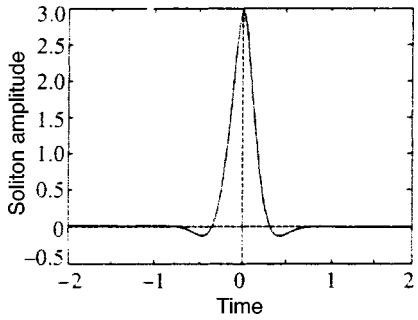

FIG. 6. Form of a soliton when $\mu/\nu_0=10$. The oscillating tail is scarcely visible.

solutions of Eq. (3.34), are simultaneously solutions in the form of stationary solitons for Eq. (2.1) with the dielectric constant

$$\varepsilon(\omega)=\varepsilon_0-a(\omega^2-\omega_0^2)^2 \quad \text{and} \quad \chi a>0.$$

As for the dispersion (3.31), here the situation is similar to what occurs for the NLSE with quadratic dispersion (3.20). Solitons are possible for $\lambda>-\nu_0^4$. The only difference from quadratic dispersion is the change in the value of the critical velocity. Near these points the structure of the solitons has the universal form (2.23).

4. STABILITY OF SOLITONS

Let us examine the stability of the solitons obtained in the preceding section. We first show how stability can be proved for the NLSE with quadratic dispersion (3.20). The Hamiltonian for it has the form

$$H=\int (|\psi_t|^2-|\psi|^4)dt\equiv I_1-I_2, \tag{4.1}$$

and the soliton solution (3.24) has the form of the stationary point of the Hamiltonian for a fixed momentum

$$P=-i\int \psi\psi_t^* dt$$

and a fixed number of particles (energy) $N=\int|\psi|^2dt$:

$$\delta(H+\beta P+\lambda N)=0.$$

Following Ref. 21, we shall prove stability in the sense of Lyapunov, i.e., we shall show that the soliton has a minimum for H at fixed P and N. For this purpose, it is convenient to represent λ in the form of a sum of $\beta^2/4$ and the positive quantity μ^2. We next consider the functional $F=H+\beta P+(\beta^2/4)N$, which, as can easily be seen, is the same Hamiltonian in a moving coordinate frame: the replacement of the wave function $\psi\to\psi e^{it\beta/2}$ transforms F into H (4.1). Thus, for stability it is sufficient to establish that H has a minimum in the stationary soliton.

Let us consider the integral $I_2=\int|\psi|^4dt$. It is easy to prove that the following chain of inequalities holds (see also Refs. 21 and 22):

$$\begin{aligned}\int_{-\infty}^{\infty}|\psi|^4dt&\leqslant\max_t|\psi|^2\int_{-\infty}^{\infty}|\psi|^2dt\\&=\int_{-\infty}^{t_{\max}}\frac{d|\psi|^2}{dt}dt\int_{-\infty}^{\infty}|\psi|^2dt\\&\leqslant 2N\int_{-\infty}^{t_{\max}}|\psi||\psi_t|dt\leqslant 2N\int_{-\infty}^{\infty}|\psi||\psi_t|dt\\&\leqslant 2N^{3/2}\left[\int_{-\infty}^{\infty}|\psi_t|^2dt\right]^{1/2}.\end{aligned} \tag{4.2}$$

This inequality can be enhanced by finding the best constant [instead of 2 in (4.2)]. The maximum value of the functional

$$G[\psi]=\frac{I_2}{N^{3/2}I_1^{1/2}}$$

clearly solves this problem. To find the maximum of $G[\psi]$ it is sufficient to consider all the stationary points of this functional and then to select the one which has the maximum value of G. All the stationary points of $G[\psi]$ are determined from the following equation, which coincides with the equation for a stationary soliton:

$$-\mu^2\psi+\psi_{tt}+2|\psi|^2\psi=0,$$

where $\lambda=\mu^2>0$. Hence it can easily be seen that the maximum of $G[\psi]$ is achieved in a real soliton solution, which is unique (to within a constant phase multiplier):

$$\psi_s=\frac{\mu}{\cosh(\mu t)}.$$

After this, all the integrals in $G[\psi]$ are easily calculated:

$$N=2\mu, \quad I_{1s}=\frac{2}{3}\mu^3, \quad I_{2s}=\frac{4}{3}\mu^3,$$

and the inequality (4.2) ultimately takes the form

$$\int_{-\infty}^{\infty}|\psi|^4dt\leqslant\frac{1}{\sqrt{3}}N^{3/2}\left[\int_{-\infty}^{\infty}|\psi_t|^2dt\right]^{1/2}. \tag{4.3}$$

The substitution of this inequality into (4.1) gives the following estimate:

$$H\geqslant H_s+(\sqrt{I_1}-\sqrt{I_{1s}})^2,$$

where $H_s=-2\mu^3/3<0$ is the value of the Hamiltonian in the soliton solution. This estimate becomes exact in the soliton solution, proving the stability of the solitons with quadratic dispersion in the sense of Lyapunov. We stress that this proof provides for the stability of solitons not only with respect to small perturbations, but also with respect to finite perturbations.

Now let us turn to fourth-order dispersion. We represent the corresponding functional $F=H+\beta P+\lambda N$ in the form of a sum of the mean value of the operator $L(i\partial_t)$ (3.29) and the nonlinear term:

$$F=\int\psi^*L(i\partial_t)\psi dt-\int|\psi|^4dt. \tag{4.4}$$

To prove the stability of solitons, we must find the analog of the inequality (4.3) for the mean $\langle L(i\partial_t)\rangle$.

Let $L(\Omega)$ be the positive definite polynomial $\Omega\in(-\infty,\infty)$ of degree $N=2l$. Then $L(\Omega)$ can be expanded as

$$L_{2l}(\Omega)=\sum_{p=0}^{l} L_{2l-2p}(\Omega_p)\prod_{i=1}^{p-1}(\Omega-\Omega_i)^2, \tag{4.5}$$

where Ω_i and the polynomials $L_{2l-2p}(\Omega)$ are constructed from L_{2l} according to the following rule. Let $\Omega=\Omega_0$ be the minimum point of $L_{2l}(\Omega)$: $\min L_{2l}(\Omega)=L_{2l}(\Omega_0)$. The latter allows us to write $L_{2l}(\Omega)$ in the form

$$L_{2l}(\Omega)=L_{2l}(\Omega_0)+(\Omega-\Omega_0)^2L_{2l-2}(\Omega),$$

where $L_{2l-2}(\Omega)$ is a nonnegative polynomial of degree $2l-2$. The expansion of the polynomial $L_{2l-2}(\Omega)$ gives a new nonnegative polynomial of degree $2l-4$. Further recursion leads us to formula (4.5). It is important that all the coefficients in this expansion are nonnegative: $L_{2l-2p}(\Omega_p)\geqslant 0$. It is also clear that $L_0(\Omega_l)=C_{2l}$.

Expansion (4.5) generates the corresponding expansion for the mean value of $L_{2l}(i\partial_t)$:

$$\langle L_{2l}(i\partial_t)\rangle\equiv\int \psi^* L_{2l}(i\partial_t)\psi dt = L_{2l}(\Omega_0)N_0+L_{2l-2}(\Omega_1)N_1+\ldots+L_0(\Omega_l)N_l, \tag{4.6}$$

where

$$N_p=\int|\psi_p|^2dt;\quad \psi_p=\prod_{q=0}^{p-1}(i\partial_t+\Omega_q)\psi,\quad p\geqslant 1;$$

$$\psi_0\equiv\psi.$$

This representation shows how the square of the norm of the positive definite polynomial operator expands in the norms N_p with the nonnegative coefficients $L_{2l-2p}(\Omega_p)$.

For the positive definite fourth-order dispersion (3.29)

$$L(\Omega)=\lambda-\beta\Omega+D\Omega^2+\gamma\Omega^3+\Omega^4$$

the expansion (4.5) reads as

$$L(\Omega)=\mu^4+\eta^2(\Omega-\Omega_0)^2+(\Omega-\Omega_0)^2(\Omega-\Omega_1)^2, \tag{4.7}$$

where μ^4 replaces $L_4(\Omega_0)$, and η^2 replaces $L_2(\Omega_1)$. With no loss of generality, we can set $\Omega_0=-\Omega_1=\nu_0$ in Eq. (4.7) (this corresponds to the replacement $\psi\to\psi\exp\{-i(\Omega_0+\Omega_1)t/2\}$), so that Eq. (4.7) takes the form

$$L(\Omega)=\mu^4+\eta^2(\Omega-\nu_0)^2+(\Omega^2-\nu_0^2)^2. \tag{4.8}$$

The difference between the dispersions (3.30) and (3.31) stems from the fact that the quantity $2\nu_0^2-\eta^2$ can be positive or negative. For (3.30) $2\nu_0^2>\eta^2$, and for (3.30) $2\nu_0^2<\eta^2$. In accordance with (4.8), the integral expansion of the norm of the operator L is written as

$$\langle L(i\partial_t)\rangle=\mu^4N+\eta^2\int|(i\partial_t+\nu_0)\psi|^2dt + \int|(\partial_t^2+\nu_0^2)\psi|^2dt. \tag{4.9}$$

This representation means that a moving soliton can be regarded as a stationary point of the new Hamiltonian

$$H'=\eta^2\int|(i\partial_t+\nu_0)\psi|^2dt+\int|(\partial_t^2+\nu_0^2)\psi|^2dt-\int|\psi|^4dt \tag{4.10}$$

when the number of particles N is fixed:

$$\delta(H'+\mu^4N)=0. \tag{4.11}$$

If the Hamiltonian H' is bounded from below for a fixed value of N, and its lower bound corresponds to a soliton, the soliton will be stable.

In terms of the new Hamiltonian the soliton solution obeys the equation

$$\mu^4\psi_s+\eta^2(i\partial_t+\nu_0)^2\psi_s+(\partial_t^2+\nu_0^2)^2\psi_s-2|\psi_s|^2\psi_s=0. \tag{4.12}$$

Next, multiplying this equation by ψ_s^* and integrating over t, we arrive at the following relation between the integrals appearing in H':

$$\mu^4N_s+\eta^2\int|(i\partial_t+\nu_0)\psi_s|^2dt+\int|(\partial_t^2+\nu_0^2)\psi_s|^2dt - 2\int|\psi_s|^4dt\equiv H_s'+\mu^4N_s-\int|\psi_s|^4dt=0.$$

Another relation follows after the multiplication of (4.12) by $t\partial_t\psi_s^*$ and integration:

$$(\mu^4+\eta^2\nu_0^2+\nu_0^4)N_s+(2\nu_0^2-\eta^2)\int|\partial_t\psi_s|^2dt - 3\int|\partial_t^2\psi_s|^2dt-\int|\psi_s|^4dt=0.$$

Combining these two relations, we obtain

$$H_s'=(\eta^2\nu_0^2+\nu_0^4)N_s+(2\nu_0^2-\eta^2)\int|\partial_t\psi_s|^2dt - 3\int|\partial_t^2\psi_s|^2dt.$$

For both dispersions the Hamiltonian H_s' is bounded from above in the soliton solution by the number of particles multiplied by a certain positive factor: for (3.30)

$$H_s'\leqslant\left[\frac{1}{12}(2\nu_0^2-\eta^2)^2+\eta^2\nu_0^2+\nu_0^4\right]N_s,$$

and for (3.31)

$$H_s'\leqslant(\eta^2\nu_0^2+\nu_0^4)N_s.$$

We now prove that H' has a lower bound for a fixed value of N. For this purpose we first evaluate the two integrals

$$J_1=\int|(i\partial_t+\nu_0)\psi|^2dt\quad\text{and}\quad J_2=\int|(\partial_t^2+\nu_0^2)\psi|^2dt$$

in terms of two other integrals: N and $I_2=\int|\psi|^4dt$. It is easy to see that the estimate (4.3) is valid for the first integral J_1:

$$\int_{-\infty}^{\infty}|\psi|^4dt\leqslant\frac{1}{\sqrt{3}}N^{3/2}\left[\int_{-\infty}^{\infty}|(i\partial_t+\nu_0)\psi|^2dt\right]^{1/2}. \tag{4.13}$$

Using the inequality (4.3) again, we can obtain the estimate sought for J_2, if we first perform integration by parts in $\int|\psi_t|^2dt$ using the Cauchy–Bulyakovskiĭ inequality,

$$\int |\psi_t|^2dt = -\int \psi^*(\psi_{tt}+\nu_0^2\psi)dt+\int \nu_0^2|\psi|^2dt \leqslant N^{1/2}\left[\int |(\partial_t^2+\nu_0^2)\psi|^2dt\right]^{1/2}+\nu_0^2N,$$

and then substitute the result obtained into (4.3):

$$J_2 \geqslant \frac{1}{N}\left(\frac{3I_2^2}{N^3}-\nu_0^2N\right)^2. \tag{4.14}$$

Using the inequalities (4.13) and (4.14) we obtain an estimate of H' in terms of N and I_2:

$$H' \geqslant f(I_2)=\frac{3I_2^2}{N^3}+\frac{1}{N}\left(\frac{3I_2^2}{N^3}-\nu_0^2N\right)^2-I_2. \tag{4.15}$$

Continuing this inequality, we obtain

$$f(I_2)\geqslant 2\frac{\sqrt{3}I_2}{N^2}\left(\frac{3I_2^2}{N^3}-\nu_0^2N\right)-I_2.$$

Finally, from this we arrive at the desired inequality, i.e., the boundedness of the Hamiltonian:

$$H' \geqslant -\frac{4\sqrt{3}N}{9}\left[1+\frac{\sqrt{3}N}{6\nu_0^2}\right]^{3/2}. \tag{4.16}$$

According to Lyapunov's theorem, this proves the stability of the stationary point of the Hamiltonian corresponding to its minimum. This minimum point is a certain soliton solution of Eq. (4.12). It need not be unique. It is noteworthy that, according to the estimate (4.16), the Hamiltonian can take negative values. If initially we have $H'<0$, the maximum value of $|\psi|^2$ will be bounded from below by the conserved quantity (compare Ref. 21):

$$\max_t |E|^2 \geqslant |H'|/N.$$

Thus, an initially existing intensity maximum cannot vanish as the pulse propagates (as x increases). On the other hand, small-amplitude radiation should ensure relaxation of the initial distribution toward a certain soliton state, which is possible owing to the lower bound on the Hamiltonian.

To conclude this section we wish to say a few words about the stability of the stationary solitons (2.23). Near the critical velocity this question can be treated within the parabolic NLSE (3.20), for which the answer is already known. As for the stability of solitons with velocities far from the critical value, the terms for dispersion of the next order must be taken into account. As we saw in this section, the fourth-order terms, which ensure that the corresponding operator L is positive, also provide for the stability of solitons. We assume that the positive definite four-order polynomial operators should ensure the stability of one-dimensional solitons. It is possible that the solitons will be unstable only for operators which increase at infinity ($|\Omega|\to\infty$) in proportion to $\sqrt{|\Omega|}$.

5. CONCLUDING REMARKS

In conclusion, we would like to note that the selection rules for solitons based on the criteria (2.8) and (2.9) are valid for arbitrary dimensionality. It is significant that the conditions for the existence of solitons remain unchanged: the corresponding operator L must be sign-definite. In addition, the fourth-order dispersion for all physical dimensionalities D ensures the existence of stable solitons for the GNLSE with cubic nonlinearity (with neglect of its dispersion). This follows from the estimate of the dispersion term of the Hamiltonian in terms of I_2 and N. In this case the inequality (4.3) has the form

$$\int |\psi|^4d^Dx \leqslant C\left[\int |\Delta\psi|^2d^Dx\right]^{D/4}\left[\int |\psi|^2d^Dx\right]^{2-D/4}. \tag{5.1}$$

Substituting this estimate into the Hamiltonian

$$H=\int |\Delta\psi|^2d^Dx-\int |\psi|^4d^Dx$$

gives its lower bound:

$$H \geqslant \int |\Delta\psi|^2d^Dx - C\left[\int |\Delta\psi|^2d^Dx\right]^{D/4}\left[\int |\psi|^2d^Dx\right]^{2-D/4} \geqslant -\left(\frac{4}{D}-1\right)\left(\frac{4}{CD}\right)^{4/(D-4)}N^{(8-D)/(4-D)}.$$

Apart from soliton stability, for media with Kerr nonlinearity this also proves that wave collapse ceases because of fourth-order dispersion for the physical dimensionalities $D=2,3$.

One last remark: in the present work we confined ourselves to consideration of equations with only cubic nonlinearity, although in the general expansion of the electric displacement D (2.2) the term which is quadratic with respect to the amplitude must be taken into account. If tangency occurs at a nonzero frequency, the quadratic anharmonic terms are not resonant near the critical velocity and can be eliminated by a canonical transformation (for further details regarding this, see the review in Ref. 14). These terms lead to renormalization of the four-wave matrix element (3.8). Thus, the universality of the behavior of solitons near the critical velocity remains in force.

We thank F. Dias for providing references pertaining to gravitational–capillary solitons. We also thank A. I. D'yachenko for performing the numerical calculations. This work was partially supported by the INTAS program and the Russian Fund for Fundamental Research (Project 97-01-00093).

*)E-mail: zakharov@itp.ac.ru
†)E-mail: kuznetso@itp.ac.ru

1)In dimensional variables the parameter β introduced here, which has the meaning of the reciprocal of the velocity, is equal to the difference between the soliton velocity and the group velocity divided by v_{gr}^2. In this case β is assumed to be small compared with $1/v_{gr}$.

[1] C. S. Gardner, J. M. Green, M. D. Kruskal, and R. B. Miura, Phys. Rev. Lett. **19**, 1095 (1967).

[2] V. E. Zakharov and A. B. Shabat, Zh. Éksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[3] A. Hasegawa and F. Tappet, Appl. Phys. Lett. **23**, 142 (1973).
[4] L. F. Mollenauer, R. H. Stolen, and M. N. Islam, Opt. Lett. **10**, 229 (1985).
[5] L. F. Mollenauer, E. Lichtman, M. J. Neibelt, and G. T. Harvey, Electron. Lett. **29**, 910 (1993).
[6] G. P. Agrawal, *Nonlinear Fiber Optics*, Academic Press, Boston (1989) [Russ. transl., Mir, Moscow (1996)].
[7] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon, Oxford (1984).
[8] S. P. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons: the Inverse Scattering Method*, Consultants Bureau, New York (1984).
[9] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform*, SIAM, Philadelphia (1985).
[10] A. C. Newell, *Solitons in Mathematics and Physics*, SIAM, Philadelphia (1985).
[11] M. S. Longuet-Higgins, J. Fluid Mech. **200**, 951 (1989); **252**, 703 (1993).
[12] G. Iooss and K. Kirchgassner, C. R. Acad. Sci., Ser. I: Math. **311**, 265 (1991).
[13] J.-M. Vanden-Broeck and F. Dias, J. Fluid Mech. **240**, 549 (1992); F. Dias and G. Iooss, Physica D **65**, 399 (1993).
[14] V. E. Zakharov and E. A. Kuznetsov, Usp. Fiz. Nauk. **167**, 1137 (1997) [Phys. Usp. **40**, 1087 (1997)].
[15] N. Bloembergen, *Nonlinear Optics*, Benjamin, Reading, Mass. (1977).
[16] J. Nycander, Chaos **4**, 253 (1994).
[17] E. M. Gromov and V. I. Talanov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **39**, 735 (1996); Zh. Éksp. Teor. Fiz. **110**, 137 (1996) [JETP **83**, 79 (1996)].
[18] R. Hirota, J. Phys. Soc. Jpn. **33**, 1456 (1973); J. Math. Phys. **14**, 805 (1973).
[19] V. K. Mezentsev and S. K. Turitsyn, Sov. J. Quantum Electron. **21**, 555 (1991); Sov. Lightwave Commun. **1**, 263 (1991).
[20] M. Klauder, W. Laedke, K. H. Spatschek, and S. K. Turitsyn, Phys. Rev. E **47**, R3844 (1993).
[21] E. A. Kuznetsov, A. N. Rubenchik, and V. E. Zakharov, Phys. Rep. **142**, 103 (1986).
[22] O. A. Ladyzhenskaya, *The Mathematical Theory of Viscous Incompressible Flow*, 2nd ed., Gordon and Breach, New York (1969).

Translated by P. Shelnitz

ERRATA

Erratum: Anomalous light absorption in molecular crystals [JETP 85, 554–564 (September 1997)]

V. I. Tovstenko

Institute of Physics, Ukrainian National Academy of Sciences, 25260 Kiev, Ukraine

Zh. Éksp. Teor. Fiz. **113**, 1915 (May 1998)

[S1063-7761(98)02505-0]

On page 555 the last term in Eq. (1) should be: $\sum_{\mathbf{nm}}{}' M^0_{\mathbf{nm}} V^+_{\mathbf{m}} V_{\mathbf{n}} B^+_{\mathbf{m}} B_{\mathbf{n}}$.

Erratum: On the nature of turbulence [JETP 86, 107–114 (January 1998)]

L. N. Pyatnitskiĭ

Institute of High Temperatures, Russian Academy of Sciences, 127412 Moscow, Russia

Zh. Éksp. Teor. Fiz. **113**, 1916 (May 1998)

[S1063-7761(98)02605-5]

On page 111, the figure should have a caption reading: FIG. 2. Structure of the flow in front of the flame.

Equations of motion of a spinning relativistic particle in external fields

A. A. Pomeranskiĭ*) and I. B. Khriplovich†)

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

(Submitted 27 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1537–1557 (May 1998)

We consider the motion of a spinning relativistic particle in external electromagnetic and gravitational fields to first order in the external field but to arbitrary order in the spin. The influence of the spin on the particle trajectory is properly accounted for by describing the spin noncovariantly. Specific calculations are performed through second order in the spin. A simple derivation is presented for the gravitational spin–orbit and spin–spin interactions of a relativistic particle. We discuss the gravimagnetic moment (GM), a particular spin effect in general relativity. We show that for a Kerr black hole the gravimagnetic ratio, i.e., the coefficient of the GM, equals unity (just as the gyromagnetic ratio equals 2 for a charged Kerr hole). The equations of motion obtained for a spinning relativistic particle in an external gravitational field differ substantially from the Papapetrou equations. © *1998 American Institute of Physics.* [S1063-7761(98)00105-X]

1. INTRODUCTION

The problem of the motion of a particle with internal angular momentum (spin) in an external field consists of two parts: the description of the spin precession and accounting for the influence of the spin on the trajectory of motion. To the lowest nonvanishing order in c^{-2}, the complete solution for the case of an external electromagnetic field was given more than 70 years ago.[1] The precession of a gyroscope in a centrally symmetric gravitational field was considered to the same approximation even earlier.[2] Spin precession for the case of the gravitational spin–spin interaction was studied much later.[3] The fully relativistic problem of spin precession in an external electromagnetic field was also solved more than 70 years ago,[4] and then in a more convenient formalism, using the covariant spin vector, in Ref. 5.

The situation with the second part of the problem, which refers to how the spin influences the trajectory, is different. The covariant equations of motion for a spinning relativistic particle in an electromagnetic field were written by Frenkel in the same paper,[4] and for the case of a gravitational field in Ref. 6. These equations have been discussed repeatedly from various points of view in many papers (see, e.g., Refs. 7–16 and 38). The problem of the influence of the spin on the trajectory of a particle in external fields is not only of theoretical interest: it is related to the problem of describing the motion of ultrarelativistic particles in accelerators[17] (see a recent review by Heinemann[18]). There are also macroscopic objects whose internal rotation affects their motion in an external gravitational field: Kerr black holes. This problem is especially important in calculations of the gravitational radiation of binary stars. In this connection it was examined in Refs. 19–22. However, when we turned to these calculations, we found[23] that even in the simpler case of an external field the equations of motion with spin taken into account to the lowest nonvanishing order in c^{-2} used in these papers lead to results that differ from the well-known gravitational spin–orbit interaction. The problem is essentially related to the proper definition of the center-of-mass coordinate. Moreover, it was found that in the same c^{-2} approximation the Papapetrou equations[6] also fail to reproduce the result for the gravitational spin–orbit interaction found in Ref. 2. This discrepancy was pointed out long ago in Ref. 24; however, the explanation suggested in that paper does not appear to be satisfactory (see Ref. 23).

In the present work we derive the equations of motion of a relativistic particle with a noncovariant description of spin. They agree with well-known limiting cases. Although for an external electromagnetic field such equations have been obtained earlier to first order in the spin[17] (see also Ref. 18), we would like to start with comments related to the use of this approximation in electrodynamics.

2. COVARIANT AND NONCOVARIANT EQUATIONS OF MOTION OF A SPINNING PARTICLE IN AN ELECTROMAGNETIC FIELD

The interaction of spin and an external electromagnetic field is described, through to terms of order c^{-2}, by the well-known Hamiltonian (see, e.g., Ref. 25)

$$H=-\frac{eg}{2m}\mathbf{s}\cdot\mathbf{B}+\frac{e(g-1)}{2m^2}\mathbf{s}\cdot(\mathbf{p}\times\mathbf{E}), \tag{1}$$

where **B** and **E** are the external magnetic and electric fields; e, m, **s**, and **p** are the particle charge, mass, spin, and momentum, and g is the gyromagnetic ratio. Note that the structure of the second (Thomas) term in this expression has not only been firmly established theoretically but has also been confirmed with high accuracy experimentally, at any rate in atomic physics. To avoid misinterpretations, we also note that the second term on the right-hand side of Eq. (1) should have been written in Hermitian form (see, e.g., Ref. 26):

$$\mathbf{p}\times\mathbf{E}\to\frac{1}{2}(\mathbf{p}\times\mathbf{E}-\mathbf{E}\times\mathbf{p})=\mathbf{p}\times\mathbf{E}+\frac{i}{2}\nabla\times\mathbf{E}.$$

But since we are interested primarily in the quasiclassical approximation, the interaction linear in spin the terms containing the derivatives of the fields are discarded. (Moreover, the correction containing $\nabla\times\mathbf{E}$ vanishes in the case of a potential electric field considered in Ref. 25).

We attempt to build a covariant equation of motion with spin, an equation that yields an expression for the force in the same approximation,

$$\mathbf{f}_m=\frac{eg}{2m}\mathbf{s}\mathbf{B},_m+\frac{e(g-1)}{2m}\left(\frac{d}{dt}[\mathbf{E}\times\mathbf{s}]_m-\mathbf{s}[\mathbf{v}\times\mathbf{E},_m]\right), \quad (2)$$

corresponding to the Hamiltonian (1) (here and below a comma with an index denotes a partial derivative). The covariant correction f^μ to the Lorentz force $eF^{\mu\nu}u_\nu$, linear in the spin tensor $S_{\mu\nu}$ and in the gradient of the electromagnetic field tensor $F_{\mu\nu,\lambda}$, may depend also on the 4-velocity u^μ. Since it satisfies $u_\mu u^\mu=1$, this correction must satisfy the condition $u_\mu f^\mu=0$. Only two independent structures meeting the last condition can be constructed from the quantities mentioned above. The first,

$$\eta^{\mu\kappa}F_{\nu\lambda,\kappa}S^{\nu\lambda}-F_{\lambda\nu,\kappa}u_\kappa S^{\lambda\nu}u^\mu,$$

reduces in the adopted approximation to

$$2\mathbf{s}(\mathbf{B},_m-[\mathbf{v}\times\mathbf{E},_m]),$$

and the second,

$$u^\lambda F_{\lambda\nu,\kappa}u^\kappa S^{\nu\mu},$$

reduces to

$$\frac{d}{dt}[\mathbf{s}\times\mathbf{E}]_m.$$

(Note that the structures with the contraction $F_{\nu\kappa,\lambda}S^{\kappa\lambda}$ reduce to these two expressions due to the Maxwell equations and the antisymmetry of $S_{\kappa\lambda}$.)

Obviously, no linear combination of these two structures can reproduce the correct expression (2) for the spin-dependent force.

In a somewhat less general form this was demonstrated in Ref. 23. There it was pointed out that the coordinate in the covariant equation does not coincide with the usual one. Therefore, to obtain the correct c^{-2} approximation to the covariant equation of motion one needs to perform an additional redefinition of the coordinate:

$$\mathbf{r}\to\mathbf{r}-\frac{1}{2m}[\mathbf{v}\times\mathbf{s}]. \quad (3)$$

In the case of 1/2 spin this redefinition is closely related to the Foldy–Wouthuysen transformation.[27] A generalization of this substitution to the case of arbitrary velocities was recently suggested by Heinemann.[18]

The correct equations of motion in an electromagnetic field that allow for spin to first order have been known for a fairly long time.[17] We recall that the initial physical definition of spin is noncovariant and refers to the proper reference frame of the particle: it is the three-dimensional vector **s** (or three-dimensional antisymmetric tensor) of the internal angular momentum defined in this reference frame. The covariant vector of spin S_μ and the covariant antisymmetric tensor $S_{\mu\nu}$ are obtained from the above entities by a Lorentz transformation. By the way, in this approach the constraints $u^\mu S_\mu=0$ and $u^\mu S_{\mu\nu}=0$ are valid identically. The precession frequency for spin **s** at an arbitrary velocity is well-known (see, e.g., Ref. 25):

$$\mathbf{\Omega}=\frac{e}{2m}\left\{(g-2)\left[\mathbf{B}-\frac{\gamma}{\gamma+1}\mathbf{v}(\mathbf{v}\cdot\mathbf{B})-\mathbf{v}\times\mathbf{E}\right]+2\left[\frac{1}{\gamma}\mathbf{B}-\frac{1}{\gamma+1}\mathbf{v}\times\mathbf{E}\right]\right\}, \quad (4)$$

where $\gamma=1/\sqrt{1-v^2}$. Naturally, the corresponding interaction Lagrangian (here the Lagrangian formulation is somewhat more convenient than the Hamiltonian one) is

$$L_{1s}=\mathbf{\Omega}\cdot\mathbf{s}=\frac{e}{2m}\mathbf{s}\left\{(g-2)\left[\mathbf{B}-\frac{\gamma}{\gamma+1}\mathbf{v}(\mathbf{v}\cdot\mathbf{B})-\mathbf{v}\times\mathbf{E}\right]+2\left[\frac{1}{\gamma}\mathbf{B}-\frac{1}{\gamma+1}\mathbf{v}\times\mathbf{E}\right]\right\}. \quad (5)$$

The equation of motion for the position has the usual form:

$$\left(\nabla-\frac{d}{dt}\nabla_{\mathbf{v}}\right)L_{\rm tot}=0, \quad (6)$$

where $L_{\rm tot}$ is the total system Lagrangian, and the equation of motion for the spin in general form is

$$\dot{\mathbf{s}}=\{L,\mathbf{s}\}, \quad (7)$$

where $\{\cdots,\cdots\}$ is the Poisson bracket, or

$$\dot{\mathbf{s}}=-i[L,\mathbf{s}] \quad (8)$$

in the quantum problem.

To conclude this section, let us discuss the following problem. It is far from obvious how meaningful these spin corrections to the equations of motion of elementary particles, say, electron or proton, are in general. According to the well-known Bohr argument (see Ref. 28), the additional Lorentz force due to the finite size of the wave packet of a charged particle and to the uncertainty relation exceeds the corresponding component of the Stern–Gerlach force. On the other hand, some time ago it was proposed that a beam of charged particles could be separated according to particle polarizations in a storage ring via the spin interaction with external fields.[29] Although this proposal is being actively discussed (see, e.g., the review by Heinemann[18]), the feasibility of such a scheme (at least in principle) is unclear.

3. EQUATIONS OF MOTION OF A SPINNING PARTICLE IN AN ELECTROMAGNETIC FIELD: GENERAL FORMALISM. EFFECTS LINEAR IN THE SPIN

In this section we develop a general approach to the derivation of the equations of motion in an external electromagnetic field to arbitrary order in the spin. We also reproduce in passing the well-known result (4).

Our approach is based on the following physically obvious reasoning. As long as we do not discuss internal excitation of a body moving in an external field, this body (even if it is macroscopic) can be considered an elementary particle with spin.

Hence the Lagrangian describing the interaction of spin and an external field can be derived from the amplitude

$$-eJ^{\mu}A_{\mu} \tag{9}$$

of elastic scattering of the particle with spin s by the vector potential A_{μ}. In view of the arguments presented at the end of Sec. 2, accounting for effects nonlinear in spin (which are of primary interest to us) can be physically meaningful only in the classical limit $s \gg 1$. It is basically this approximation that is used throughout the paper.

The matrix element J_{μ} of the electromagnetic current operator between states with momenta k and k' can be written (provided that P and T invariance apply) as follows (see Refs. 30 and 31):

$$J_{\mu}=\frac{1}{2\epsilon}\bar{\psi}(k')\{p_{\mu}F_e+\Sigma_{\mu\nu}q^{\nu}F_m\}\psi(k), \tag{10}$$

where $p_{\mu}=(k'+k)_{\mu}$ and $q_{\mu}=(k'-k)_{\mu}$.

The wave function ψ of a particle with arbitrary spin can be written as (see, e.g., Ref. 25, § 31)

$$\psi=\frac{1}{\sqrt{2}}\begin{pmatrix}\xi\\ \eta\end{pmatrix}. \tag{11}$$

Both spinors, $\xi=\{\xi^{\alpha_1\alpha_2\cdots\alpha_p}_{\dot\beta_1\dot\beta_2\cdots\dot\beta_q}\}$ and $\eta=\{\eta^{\beta_1\beta_2\cdots\beta_q}_{\dot\alpha_1\dot\alpha_2\cdots\dot\alpha_p}\}$, are symmetric in the dotted and undotted indices separately, and $p+q=2s$. For a particle with half-integer spin we can select $p=s+1/2$ and $q=s-1/2$. In the case of integer spin it is convenient to use $p=q=s$. The spinors ξ and η are selected so that under reflection they transform into each other (to within a phase). For $p\neq q$ they are different objects and belong to different representations of the Lorentz group. For $p=q$ the two spinors coincide. Nevertheless, we use the same expression (11) for the wave function of any spin, i.e., we also formally introduce the object η for integer spin, bearing in mind that it can be expressed in terms of ξ. This makes it possible to do calculations for integer and half-integer spins in a unified way.

In the rest frame both ξ and η coincide with the nonrelativistic spinor ξ_0, which is symmetric in all indices; in this frame there is no difference between dotted and undotted indices. The spinors ξ and η can be obtained from ξ_0 via the Lorentz transformation

$$\xi=\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}\xi_0, \quad \eta=\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}\xi_0. \tag{12}$$

Here the vector $\boldsymbol{\phi}$ is directed along the velocity, $\tanh\phi=v$,

$$\boldsymbol{\Sigma}=\sum_{i=1}^{p}\boldsymbol{\sigma}_i-\sum_{i=p+1}^{p+q}\boldsymbol{\sigma}_i,$$

and $\boldsymbol{\sigma}_i$ acts on the ith index of the spinor ξ_0 as follows:

$$\boldsymbol{\sigma}_i\xi_0=(\boldsymbol{\sigma}_i)_{\alpha_i\beta_i}(\xi_0)_{\cdots\beta_i\cdots}. \tag{13}$$

In the Lorentz transformation (12) for ξ, after the operator $\boldsymbol{\Sigma}$ has acted on ξ_0 the first p indices become identical to the upper undotted indices and the next q indices become identical to the lower dotted indices. For η the situation is just the opposite.

We note that in an external field the components of velocity $\mathbf{v}$ (and the components of $\boldsymbol{\phi}$) do not commute, in general. However, in the adopted approximation, linear in the external field, the noncommutativity, proportional to the field, can be ignored. Moreover, we are mainly interested in the classical limit of the final result, where such commutators are negligible, since they are proportional to an extra power of $\hbar$. Therefore, $\mathbf{v}$ and $\boldsymbol{\phi}$ will be treated as ordinary numerical parameters.

Next,

$$\bar{\psi}=\psi^{\dagger}\gamma_0=\psi^{\dagger}\begin{pmatrix}0 & I\\ I & 0\end{pmatrix},$$

where I is the sum of 2×2 identity matrices acting on all indices of the spinors ξ and η. The components of the matrix $\Sigma_{\mu\nu}=-\Sigma_{\nu\mu}$ are

$$\Sigma_{0n}=\begin{pmatrix}-\Sigma_n & 0\\ 0 & \Sigma_n\end{pmatrix}, \tag{14}$$

$$\Sigma_{mn}=-2i\epsilon_{mnk}\begin{pmatrix}s_k & 0\\ 0 & s_k\end{pmatrix}, \tag{15}$$

$$\mathbf{s}=\frac{1}{2}\sum_{i=1}^{2s}\boldsymbol{\sigma}_i.$$

The scalar operators $F_{e,m}$ depend on two invariants, $t=q^2$ and $\tau=(S^{\mu}q_{\mu})^2$. The covariant spin vector S_{μ} is defined, e.g., for the state with momentum k_{μ} and can be obtained via the Lorentz transformation from the spin vector $(0,\mathbf{s})$ in the rest frame:

$$S^{\mu}=(S_0,\mathbf{S}), \quad S_0=\frac{(\mathbf{s}\cdot\mathbf{k})}{m}, \quad \mathbf{S}=\mathbf{s}+\frac{\mathbf{k}(\mathbf{k}\cdot\mathbf{s})}{m(\epsilon+m)}. \tag{16}$$

In the expansion in the electric multipoles

$$F_e(t,\tau)=\sum_{n=0}^{N_e}f_{e,2n}(t)\,\tau^n,$$

the highest power N_e obviously is s and $s-1/2$ for integer and half-integer spin, respectively. In the magnetic multipole expansion

$$F_m(t,\tau)=\sum_{n=0}^{N_m}f_{m,2n}(t)\,\tau^n,$$

the highest power N_m is $s-1$ and $s-1/2$ for integer and half-integer spin, respectively. Clearly,

$$f_{e,0}(0)=1, \quad f_{m,0}(0)=\frac{g}{2}.$$

Note that we have chosen the noncovariant normalization for the amplitude (9), since we are interested in the Lagrangian referring to the world time t and not to the proper time τ.

Let us now reproduce in this approach the well-known result (5) for the case of a constant external field. We start with terms proportional to the g-factor. The corresponding term in the scattering amplitude can be written as

$$\frac{eg}{4\epsilon}\xi_0'^{\dagger}\Big\{[\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}(\mathbf{s}\cdot\mathbf{B})\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}$$
$$+\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}(\mathbf{s}\cdot\mathbf{B})\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}]+\frac{i}{2}[\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}$$
$$\times(\boldsymbol{\Sigma}\cdot\mathbf{E})\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}-\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}(\boldsymbol{\Sigma}\cdot\mathbf{E})$$
$$\times\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}]\Big\}\xi_0. \tag{17}$$

What is important in the case of a constant external field is that we can put $\mathbf{k}'=\mathbf{k}$, $\mathbf{v}'=\mathbf{v}$, and $\boldsymbol{\phi}'=\boldsymbol{\phi}$, since $\mathbf{q}=\mathbf{k}'-\mathbf{k}$ corresponds to the field gradient.

In our further calculations we use the well-known identity

$$\exp\{\hat{A}\}\hat{B}\,\exp\{-\hat{A}\}=\hat{B}+\frac{1}{1!}[\hat{A},\hat{B}]$$
$$+\frac{1}{2!}[\hat{A},[\hat{A},\hat{B}]]+\cdots,$$

and the following relationships:

$$[\Sigma_i,\Sigma_j]=4i\epsilon_{ijk}s_k,\quad [\Sigma_i,s_j]=i\epsilon_{ijk}\Sigma_k, \tag{18}$$

$$\cosh\phi=\gamma,\quad \sinh\phi=v\gamma. \tag{19}$$

After some simple algebraic transformations, the expression (17) reduces to

$$\frac{eg}{2m}\mathbf{s}\cdot\left[\mathbf{B}-\frac{\gamma}{\gamma+1}\mathbf{v}(\mathbf{v}\cdot\mathbf{B})-\mathbf{v}\times\mathbf{E}\right]. \tag{20}$$

Let us now discuss the contribution of the convection term

$$-\frac{e}{2\epsilon}\overline{\psi}(k')\psi(k)p^{\mu}A_{\mu}. \tag{21}$$

We write the product of the exponentials in

$$\overline{\psi}(k')\psi(k)=\tfrac{1}{2}\xi_0'^{\dagger}[\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}'/2\}\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}$$
$$+\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}'/2\}\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}]\xi_0 \tag{22}$$

as

$$\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}'/2\}\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}$$
$$=\prod_p\,\exp\{\boldsymbol{\sigma}\cdot\boldsymbol{\phi}'/2\}\exp\{-\boldsymbol{\sigma}\cdot\boldsymbol{\phi}/2\}\prod_q\,\exp\{-\boldsymbol{\sigma}\cdot\boldsymbol{\phi}'/2\}$$
$$\times\exp\{\boldsymbol{\sigma}\cdot\boldsymbol{\phi}/2\}. \tag{23}$$

Let us consider a typical factor in this formula:

$$\exp\{\boldsymbol{\sigma}\cdot\boldsymbol{\phi}'/2\}\exp\{-\boldsymbol{\sigma}\cdot\boldsymbol{\phi}/2\}$$
$$=\cosh(\phi'/2)\cosh(\phi/2)-(\mathbf{n}'\cdot\mathbf{n})\sinh(\phi'/2)$$
$$\times\sinh(\phi/2)+\boldsymbol{\sigma}\cdot[\mathbf{n}'\sinh(\phi'/2)\cosh(\phi/2)$$
$$-\mathbf{n}\,\cosh(\phi'/2)\sinh(\phi/2)]$$
$$-i(\boldsymbol{\sigma}\cdot[\mathbf{n}'\times\mathbf{n}])\sinh(\phi'/2)\sinh(\phi/2), \tag{24}$$

where $\mathbf{n}'=\mathbf{v}'/v'$, and $\mathbf{n}=\mathbf{v}/v$. Here we are interested in gradients only as long as they enter into the expression together with the spin, in $\cosh(\phi'/2)\cosh(\phi/2)-(\mathbf{n}'\cdot\mathbf{n})\times\sinh(\phi'/2)\times\sinh(\phi/2)$ we put $\phi'=\phi$ and $\mathbf{n}'=\mathbf{n}$, after which this combination becomes equal to unity. Since we are interested in an interaction which is linear in the spin, the product (23) reduces to

$$1+\boldsymbol{\Sigma}\cdot[\mathbf{n}'\sinh(\phi'/2)\cosh(\phi/2)-\mathbf{n}\,\cosh(\phi'/2)\sinh(\phi/2)]$$
$$-2i(\mathbf{s}\cdot[\mathbf{n}'\times\mathbf{n}])\sinh(\phi'/2)\sinh(\phi/2).$$

When this combination is plugged into (22), the terms proportional to $\boldsymbol{\Sigma}$ cancel out. If we now limit ourselves to terms linear in $\mathbf{q}$, we reduce the spin-dependent part of (21) to

$$-e\frac{p^{\mu}}{2\epsilon}\frac{i(\mathbf{s}\cdot\mathbf{k}\times\mathbf{q})}{m(\epsilon+m)}A_{\mu}.$$

We note that since $p^{\mu}q_{\mu}=0$, the following identity holds:

$$p^{\mu}q_{\alpha}A_{\mu}=p^{\mu}(q_{\alpha}A_{\mu}-q_{\mu}A_{\alpha})=p^{\mu}iF_{\alpha\mu}. \tag{25}$$

Now we can put $p_{\mu}\to 2mu_{\mu}$, where u_{μ} is the 4-velocity. As a result we arrive at the following expression:

$$-\frac{e}{2m}\mathbf{s}\cdot\left[2\left(1-\frac{1}{\gamma}\right)\mathbf{B}-\frac{2\gamma}{\gamma+1}\mathbf{v}(\mathbf{v}\cdot\mathbf{B})-\frac{2\gamma}{\gamma+1}\mathbf{v}\times\mathbf{E}\right]. \tag{26}$$

The sum of (20) and (26) yields (5). Thus, we have reproduced the well-known result for the interaction linear in the spin, starting from the relativistic wave equation for arbitrary spin.

Below we repeatedly use identities of the form (25). In classical terms such a transformation amounts to discarding a total time derivative in the Lagrangian (or adding such a derivative to the Lagrangian). Indeed,

$$u^{\mu}q_{\mu}\to u^{\mu}\partial_{\mu}=\gamma\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla\right)=\gamma\frac{d}{dt}.$$

4. EQUATIONS OF MOTION OF A SPINNING PARTICLE IN AN ELECTROMAGNETIC FIELD. EFFECTS QUADRATIC IN THE SPIN

Let us now investigate the interaction of second order in the spin. The explicitly quadrupole ''bare'' interaction in (9) and (10) is

$$-e\frac{p^{\mu}}{2\epsilon}f_{e,2}(S^{\alpha}q_{\alpha})^2A_{\mu}. \tag{27}$$

Using the identity (27) and Eqs. (16), we reduce the interaction to

$$-ef_{e,2}\gamma\left\{(\mathbf{v}\cdot\mathbf{s})\left[\frac{\partial}{\partial t}+\frac{\gamma}{\gamma+1}(\mathbf{v}\cdot\nabla)\right]+\frac{1}{\gamma}\mathbf{s}\cdot\nabla\right\}\left[(\mathbf{s}\cdot\mathbf{E})\right.$$
$$\left.-\frac{\gamma}{\gamma+1}(\mathbf{s}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{E})+\mathbf{s}\cdot\mathbf{v}\times\mathbf{B}\right].$$

Discarding the total time derivative $\partial/\partial t+\mathbf{v}\cdot\boldsymbol{\nabla}$, we arrive at the following expression:

$$L_{2s}=-ef_{e,2}\left[\mathbf{s}\cdot\nabla-\frac{\gamma}{\gamma+1}(\mathbf{v}\cdot\mathbf{s})(\mathbf{v}\cdot\nabla)\right]\left[\mathbf{s}\cdot\mathbf{E}-\frac{\gamma}{\gamma+1}\right.$$
$$\left.\times(\mathbf{s}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{E})+\mathbf{s}\cdot(\mathbf{v}\times\mathbf{B})\right]. \tag{28}$$

If we use the Maxwell equations and add a total derivative with respect to t, we can show that the tensor s_is_j in (28) can be written in the following irreducible form:

$$s_is_j\to s_is_j-(1/3)\,\delta_{ij}\mathbf{s}^2.$$

Now, from the nonrelativistic limit of (28) it follows that (28) indeed describes the interaction between the external field and the quadrupole moment

$$Q_{ij}=-2ef_{e,2}(3s_is_j-\delta_{ij}\mathbf{s}^2), \tag{29}$$
$$Q=Q_{zz}|_{s_z=s}=-2ef_{e,2}s(2s-1).$$

As $\gamma\to\infty$, the interaction (28) tends to the constant

$$L_{2s}=-ef_{e,2}[\mathbf{s}\cdot\nabla-(\mathbf{v}\cdot\mathbf{s})(\mathbf{v}\cdot\nabla)][\mathbf{s}\cdot\mathbf{E}-(\mathbf{s}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{E})$$
$$+\mathbf{s}\cdot(\mathbf{v}\times\mathbf{B})]. \tag{30}$$

It is well known that even in the absence of a bare quadrupole term, i.e., for $f_{e,2}=0$, a quadrupole interaction in the nonrelativistic limit arises due to the convection and magnetic terms in (9). The value of this induced quadrupole moment for a particle with arbitrary spin was obtained in Ref. 31:[1)]

$$Q_1=-e(g-1)\left(\frac{\hbar}{mc}\right)^2\begin{cases}s, & \text{integer spin},\\ s-1/2, & \text{half-integer spin}.\end{cases} \tag{31}$$

Here we have explicitly displayed the Planck constant $\hbar$ to show that the induced quadrupole moment Q_1 vanishes in the classical limit $\hbar\to0$, $s\to\infty$, $\hbar s\to$const. Hence the interaction that is second order in the spin and proportional to Q_1 has no real effect on the equations of motion of a classical particle (although it does play a role in atomic spectroscopy[31]).

The convection and magnetic terms in (9) induce an interaction of second order in the spin with a classical limit, an interaction of interest for our problem. It is convenient here to start with the convection current interaction. Let us go back to Eq. (24). In it we again put

$$\cosh(\phi'/2)\cosh(\phi/2)-(\mathbf{n}'\cdot\mathbf{n})\sinh(\phi'/2)\sinh(\phi/2)=1.$$

In the other terms which are linear in $\boldsymbol{\sigma}$ we keep only the first power of $\mathbf{q}\to-i\hbar\nabla$ in the hope that in the final result (23) $\hbar$ will be present in the combination $\hbar s\to$const. Nevertheless, these terms by themselves are small compared to unity, so that in the classical limit the expression (24) can be written as

$$\exp\{\boldsymbol{\sigma}\cdot[\mathbf{n}'\sinh(\phi'/2)\cosh(\phi/2)-\mathbf{n}\cosh(\phi'/2)\sinh(\phi/2)]$$
$$-i(\boldsymbol{\sigma}\cdot[\mathbf{n}'\times\mathbf{n}]\sinh^2(\phi/2)\}.$$

Clearly, in the product (23) the operators $\boldsymbol{\sigma}$ attached to $\mathbf{n}'\sinh(\phi'/2)\cosh(\phi/2)-\mathbf{n}\cosh(\phi'/2)\sinh(\phi/2)$ combine in the exponent of the resulting exponential into the operator $\boldsymbol{\Sigma}$, which vanishes in the classical limit. In this limit the only operators $\boldsymbol{\sigma}$ that survive are those that are attached to $(\mathbf{n}'\times\mathbf{n})\sinh^2(\phi/2)$; they combine into $2\mathbf{s}$. Thus, in the classical limit with the second identity in (19) taken into account, the product (23) reduces to

$$\exp\left\{\frac{\gamma}{\gamma+1}(\mathbf{s}\cdot\mathbf{v}\times\boldsymbol{\nabla})\right\}. \tag{32}$$

Note that the action of the operator (32) on any function of position, whether it is the vector potential or the field strength, amounts to shifting the function's argument:

$$\mathbf{r}\to\mathbf{r}+\frac{1}{m}\frac{\gamma}{\gamma+1}\mathbf{s}\times\mathbf{v}.$$

Interestingly, this substitution was suggested by Heinemann[18] for a transition from covariant equations linear in spin to noncovariant equations. Equation (3) is a particular case of this substitution in the c^{-2} approximation.

Now, taking into account the second term in the expansion of the exponential function (32) and using the identity (25) once more, we can easily obtain the following expression for the interaction quadratic in the spin that arises from the convection current:

$$-\frac{e}{2m^2}\frac{\gamma}{\gamma+1}(\mathbf{s}\cdot\mathbf{v}\times\nabla)\left[\left(1-\frac{1}{\gamma}\right)\mathbf{s}\cdot\mathbf{B}-\frac{\gamma}{\gamma+1}\right.$$
$$\left.\times(\mathbf{s}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{B})-\frac{\gamma}{\gamma+1}(\mathbf{s}\cdot\mathbf{v}\times\mathbf{E})\right]. \tag{33}$$

Now we discuss the contribution to the effect being discussed due to the magnetic moment. It is convenient to write the term in the scattering amplitude we are interested in (it is proportional to the g-factor) as

$$\frac{eg}{4\epsilon}\xi_0'^{\dagger}\left\{[\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}'/2\}\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}(\mathbf{s}\cdot\mathbf{B})\right.$$
$$\times\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}+\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}'/2\}\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}$$
$$\times\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}(\mathbf{s}\cdot\mathbf{B})\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}]+(i/2)$$
$$\times[\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}'/2\}\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}(\boldsymbol{\Sigma}\cdot\mathbf{E})$$
$$\times\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}-\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}'/2\}\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}$$
$$\left.\times\exp\{-\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}(\boldsymbol{\Sigma}\cdot\mathbf{E})\exp\{\boldsymbol{\Sigma}\cdot\boldsymbol{\phi}/2\}]\right\}\xi_0. \tag{34}$$

Using in this case the first term in the expansion of the exponential function (32), we arrive at the following expression for the contribution proportional to the magnetic moment:

$$\frac{eg}{2m^2}\frac{\gamma}{\gamma+1}(\mathbf{s}\cdot\mathbf{v}\times\nabla)\Bigg[(\mathbf{s}\cdot\mathbf{B})-\frac{\gamma}{\gamma+1}(\mathbf{s}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{B}) -(\mathbf{s}\cdot\mathbf{v}\times\mathbf{E})\Bigg]. \tag{35}$$

The final result for the induced interaction quadratic in the spin is

$$L^i_{2s}=\frac{e}{2m^2}\frac{\gamma}{\gamma+1}(\mathbf{s}\cdot\mathbf{v}\times\nabla)\Bigg[\left(g-1+\frac{1}{\gamma}\right)(\mathbf{s}\cdot\mathbf{B})-(g-1)\frac{\gamma}{\gamma+1}(\mathbf{s}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{B})-\left(g-\frac{\gamma}{\gamma+1}\right)(\mathbf{s}\cdot\mathbf{v}\times\mathbf{E})\Bigg]. \tag{36}$$

We note that in the nonrelativistic limit the induced interaction with the magnetic field tends to zero as v/c and that with an electric field, as $(v/c)^2$. Moreover, the part of the interaction (36) that is not related to the g-factor is not irreducible in the spin; in other words, $s_i s_j$ in it cannot be reduced to the irreducible tensor $s_i s_j-(1/3)\delta_{ij}\mathbf{s}^2$. In fact, the interaction (36) is not a quadrupole one. However, its asymptotic behavior for $\gamma\to\infty$ is of interest. In this limit

$$L^i_{2s}=\frac{e}{2m^2}(g-1)(\mathbf{s}\cdot\mathbf{v}\times\boldsymbol{\nabla})(\mathbf{s}\cdot\mathbf{B})-(\mathbf{s}\cdot\mathbf{v})\times(\mathbf{v}\cdot\mathbf{B})-(\mathbf{s}\cdot\mathbf{v}\times\mathbf{E}). \tag{37}$$

Surprisingly, the asymptotic formulas (30) and (37) coincide to within a factor and a total time derivative. To prove this, it is convenient to introduce a triple of orthogonal unit vectors:

$$\mathbf{v};\quad \boldsymbol{\rho}=\frac{\mathbf{v}\times\mathbf{s}}{|\mathbf{v}\times\mathbf{s}|},\quad \boldsymbol{\zeta}=\mathbf{v}\times\boldsymbol{\rho}.$$

Using the fact that this triple forms a complete basis and the equation $\dot{\mathbf{E}}=\boldsymbol{\nabla}\times\mathbf{B}$ and discarding the total derivative with respect to t, we can easily verify that

$$[(\mathbf{s}\cdot\nabla)-(\mathbf{v}\cdot\mathbf{s})(\mathbf{v}\cdot\nabla)][(\mathbf{s}\cdot\mathbf{E})-(\mathbf{s}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{E})+(\mathbf{s}\cdot\mathbf{v}\times\mathbf{B})]$$
$$=[\mathbf{v}\times\mathbf{s}]^2(\boldsymbol{\zeta}\cdot\nabla)[(\boldsymbol{\zeta}\cdot\mathbf{E})+(\boldsymbol{\rho}\cdot\mathbf{B})]$$

coincides with

$$[\mathbf{s}\cdot(\mathbf{v}\times\nabla)][(\mathbf{s}\cdot\mathbf{B})-(\mathbf{s}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{B})-\mathbf{s}\cdot(\mathbf{v}\times\mathbf{E})]$$
$$=-(\mathbf{v}\times\mathbf{s})^2(\boldsymbol{\rho}\cdot\nabla)[\boldsymbol{\rho}\cdot(\mathbf{v}\times\mathbf{B})+\boldsymbol{\rho}\cdot\mathbf{E}].$$

Thus, there is a special value of the bare quadrupole moment,

$$Q=-2(g-1)\frac{es^2}{m^2},\ \text{or}\ f_{e,2}=(g-1)\frac{1}{2m^2} \tag{38}$$

(recall that now we are considering a classical situation, where $s\gg 1$), for which the total interaction, quadratic in spin, $L_{2s}+L^i_{2s}$, asymptotically decreases with increasing energy.

The situation is similar to the one involving an interaction that is linear in spin. It is well known (see, e.g., Refs. 11, 32, and 33) that there exists a special value of the g-factor, $g=2$, at which the interaction linear in spin decreases as $\gamma\to\infty$. This follows immediately from the formula (5) for the first-order Lagrangian. Thus, if we also assume $g=2$, we obtain

$$Q=-2\frac{es^2}{m^2},\ \text{or}\ f_{e,2}=\frac{1}{2m^2}. \tag{39}$$

Note that $g=2$ for the bare magnetic moment is a necessary (but not sufficient) condition for quantum electrodynamics to be renormalizable.[11,32,33] The condition is satisfied not only for the electron but also for the charged vector boson in the renormalizable electroweak theory.

In one respect, however, the situation with the special values (38) and (39) of the quadrupole moment differs from the situation with the g-factor. The conditions (38) and (39), as distinct from the condition $g=2$, are not universal, since they are valid only for large spins, $s\gg 1$; in other words, they refer only to classical objects with internal angular momentum. In particular, for the charged vector boson of the renormalizable electroweak theory the bare quadrupole interaction is absent, $f_{e,2}=0$. The quadrupole moment of this particle is (in our terms) induced and is given by Eq. (31) at $s=1$ and $g=2$.

5. SPIN PRECESSION IN A GRAVITATIONAL FIELD

In this section we present a simple and general derivation of the equations of spin precession in a gravitational field. This approach not only makes it possible to easily reproduce and generalize known results for spin effects. The remarkable analogy detailed above between gravitational and electromagnetic fields makes it possible to easily apply the results of Secs. 3 and 4 to the case of an external gravitational field.

The law of angular momentum conservation in flat space–time and the equivalence principle imply that the spin 4-vector S^μ is in translational motion along the particle's worldline. The parallel transport of a vector along a geodesic $x^\mu(\tau)$ means that its covariant derivative vanishes:

$$\frac{DS^\mu}{D\tau}=0. \tag{40}$$

(In this section we restrict our discussion to effects that are linear in the spin.) At this point in our discussion we introduce the tetrad formalism, which is a natural way to describe spin. In view of Eq. (40), the equation for the tetrad components of spin $S^a=S^\mu e^a_\mu$ is

$$\frac{DS^a}{D\tau}=\frac{dS^a}{d\tau}=S^\mu e^a_{\mu;\nu}u^\nu=\eta^{ab}\gamma_{bcd}u^d S^c, \tag{41}$$

where

$$\gamma_{abc}=e_{a\mu;\nu}e^\mu_b e^\nu_c=-\gamma_{bac} \tag{42}$$

are the Ricci rotation coefficients.[34] Of course, the equation for the tetrad 4-velocity components is exactly the same:

$$\frac{du^a}{d\tau}=\eta^{ab}\gamma_{bcd}u^d u^c. \tag{43}$$

The meaning of Eqs. (41) and (43) is clear: the tetrad components of both vectors vary in the same way due only to the rotation of the local Lorentzian 4-frame.

In exactly the same way, the 4-dimensional spin and velocity of a charged particle with the gyromagnetic ratio $g=2$ precess with the same angular velocity in an external electromagnetic field by virtue of the Bargman–Michel–Telegdi equation[5,25] (at $g=2$) and the Lorentz equation:

$$\frac{dS_a}{dt}=\frac{e}{m}F_{ab}S^b, \quad \frac{du_a}{dt}=\frac{e}{m}F_{ab}u^b.$$

Thus we have the evident correspondence

$$\frac{e}{m}F_{ab}\leftrightarrow\gamma_{abc}u^c. \tag{44}$$

The correspondence makes it possible to obtain the precession frequency $\boldsymbol{\omega}$ of the three-dimensional spin vector $\mathbf{s}$ in an external gravitational field from (4) via the simple substitution

$$\frac{e}{m}B_i\rightarrow-\frac{1}{2}\epsilon_{ikl}\gamma_{klc}u^c, \quad \frac{e}{m}E_i\rightarrow\gamma_{0ic}u^c. \tag{45}$$

This frequency is

$$\omega_i=-\epsilon_{ikl}\left(\frac{1}{2}\gamma_{klc}+\frac{u^k}{u^0+1}\gamma_{0lc}\right)\frac{u^c}{u^0_w}. \tag{46}$$

The common factor $1/u^0_w$ in this expression is related to the passage in the left-hand side of Eq. (41) to differentiation with respect to the world time t:

$$\frac{d}{d\tau}=\frac{dt}{d\tau}\frac{d}{dt}=u^0_w\frac{d}{dt}.$$

We have attached a subscript w to the quantity u^0_w to emphasize that this is a world component of 4-velocity rather than a tetrad component. The other indices in (46) are tetrad indices, $c=0,1,2,3$ and $i,k,l=1,2,3$. The corresponding spin-dependent correction to the Lagrangian is

$$L_{1sg}=\mathbf{s}\cdot\boldsymbol{\omega}. \tag{47}$$

As an illustration of Eqs. (46) and (47), we apply them to the cases of spin–orbit and spin–spin interactions. We restrict discussion (as is common in such problems) to the linear approximation in the gravitational field. However, in our approach, in contrast to the standard approaches, both problems can easily be solved for arbitrary particle velocities.

The tetrads $e_{a\mu}$ are related to the metric as follows:

$$e_{a\mu}e_{b\nu}\eta^{ab}=g_{\mu\nu}.$$

In the linear approximation we can put $g_{\mu\nu}=\eta_{\mu\nu}+h_{\mu\nu}$, so that there is no need to distinguish between tetrad and world indices in $e_{a\mu}$. The ambiguity in the choice of tetrads can be resolved by selecting the symmetric gauge $e_{\mu\nu}=e_{\nu\mu}$. Then

$$e_{\mu\nu}=\eta_{\mu\nu}+\tfrac{1}{2}h_{\mu\nu}.$$

Using the expression (42) for the Ricci coefficients, in the linear approximation we find that

$$\gamma_{abc}=\tfrac{1}{2}(h_{bc,a}-h_{ac,b}). \tag{48}$$

We start with the spin–orbit interaction. In the centrally symmetric field created by a mass M the metric is

$$h_{00}=-\frac{2kM}{r}, \quad h_{mn}=-\frac{2kM}{r}\delta_{mn}. \tag{49}$$

Here the nonvanishing Ricci coefficients are

$$\gamma_{ijk}=\frac{kM}{r^3}(\delta_{jk}r_i-\delta_{ik}r_j), \quad \gamma_{0i0}=-\frac{kM}{r^3}r_i. \tag{50}$$

Plugging these coefficients into (46) yields the following expression for the precession frequency:

$$\boldsymbol{\omega}_{ls}=\frac{2\gamma+1}{\gamma+1}\frac{kM}{r^3}\mathbf{v}\times\mathbf{r}. \tag{51}$$

In the limit of low velocities, $\gamma\rightarrow1$, the answer goes over into the classical result.[2]

Now we examine spin–spin interaction. Let the spin of the central body be $\mathbf{s}_0$. The components of the metric that are linear in $\mathbf{s}_0$, which are responsible for the spin–spin interaction, are

$$h_{0i}=2k\frac{[\mathbf{s}_0\times\mathbf{r}]_i}{r^3}.$$

Here the nonvanishing Ricci coefficients are

$$\gamma_{ij0}=k\left(\nabla_i\frac{[\mathbf{s}_0\times\mathbf{r}]_j}{r^3}-\nabla_j\frac{[\mathbf{s}_0\times\mathbf{r}]_i}{r^3}\right), \tag{52}$$

$$\gamma_{0ij}=-k\nabla_i\frac{[\mathbf{s}_0\times\mathbf{r}]_j}{r^3}.$$

The the spin–spin precession frequency is

$$\boldsymbol{\omega}_{ss}=-k\left(2-\frac{1}{\gamma}\right)(\mathbf{s}_0\cdot\nabla)\nabla\frac{1}{r}+k\frac{\gamma}{\gamma+1}[\mathbf{v}(\mathbf{s}_0\cdot\nabla)-\mathbf{s}_0(\mathbf{v}\cdot\nabla)$$
$$+(\mathbf{v}\cdot\mathbf{s}_0)\nabla](\mathbf{v}\cdot\nabla)\frac{1}{r}. \tag{53}$$

In the low-velocity limit this formula also becomes the corresponding classical result.[3]

To conclude this section we note that in the case of an external gravitational field there is no covariant expression for the force linear in the particle spin. In other words, the deviation of the trajectory of a spinning particle from the geodesics is not described by the Riemann tensor. In this case the possible covariant structure is unique to within a factor (in Ref. 6 this factor equals $-1/2m$): $R_{\mu\nu ab}u^\nu S^{ab}$. As mentioned in the Introduction, this covariant description (as distinct from our formulas (46) and (47)) contradicts the classical results in the low-velocity limit.

6. EQUATIONS OF MOTION OF A SPINNING PARTICLE IN A GRAVITATIONAL FIELD: GENERAL APPROACH

The equations of motion in an external gravitational field to any order in spin are constructed similarly to the equations of motion in the case of an electromagnetic field.

We start the elastic scattering amplitude in a weak external gravitational field $h_{\mu\nu}$. This amplitude will be used only as a guide, and later we will step outside the linear approximation. The amplitude is

$$-\tfrac{1}{2}T_{\mu\nu}h^{\mu\nu}. \tag{54}$$

The matrix element $T_{\mu\nu}$ of the energy–momentum tensor between states with momenta k and k' can be written as

$$T_{\mu\nu}=\frac{1}{4\epsilon}\,\overline{\psi}(k')\Big\{p_\mu p_\nu F_1+\frac{1}{2}(p_\mu\Sigma_{\nu\lambda}+p_\nu\Sigma_{\mu\lambda})q^\lambda F_2 +(\eta_{\mu\nu}q^2-q_\mu q_\nu)F_3+[S_\mu S_\nu q^2-(S_\mu q_\nu+S_\nu q_\mu) \times(Sq)+\eta_{\mu\nu}(Sq)^2]F_4\Big\}\psi(k). \tag{55}$$

The scalar operators F_i in this expression are also expanded in powers of $\tau=(Sq)^2$:

$$F_i(t,\tau)=\sum_{n=0}^{N_i} f_{i,2n}(t)\tau^n.$$

Clearly, the total number of invariant form factors $f_{i,2n}$ is $4s+2$ and $4s+1$ for integer and half-integer spin, respectively. The independence of the four tensor structures in (55) is obvious. As for the completeness of the expansion, it can be verified, e.g., by calculating the total number of invariant form factors in the annihilation channel and showing that it coincides with the above result.

In the generally covariant notation, the structure $(\eta_{\mu\nu}q^2-q_\mu q_\nu)h^{\mu\nu}$ corresponds to the scalar curvature R, and $[S_\mu S_\nu q^2-(S_\mu q_\nu+S_\nu q_\mu)(Sq)+\eta_{\mu\nu}(Sq)^2]h^{\mu\nu}$ corresponds to the product $R_{\mu\nu}S^\mu S^\nu$, where $R_{\mu\nu}$ is the Ricci tensor. Since we are interested in the equations of motion in a sourceless field, the corresponding terms in the expansion (55) are omitted.

Just as in electrodynamics charge conservation leads to the condition $f_{e,0}(0)=1$, here energy conservation leads to $f_{1,0}(0)=1$. As for the term in the amplitude (54) containing $f_{2,0}$, it is convenient to write it in a different form, using the analogy (44) with an electromagnetic field. Putting $g=2$ and $(e/m)F_{ab}=f_{ab}=\gamma_{abc}u^c$ in the corresponding electromagnetic term

$$i\frac{eg}{8\epsilon}\,\overline{\psi}(k')\Sigma^{ab}F_{ab}\psi(k),$$

we arrive at the following contribution to the Lagrangian of the gravitational interaction:

$$i\,\frac{1}{4u_w^0}\,\overline{\psi}(k')\Sigma^{ab}f_{ab}\psi(k), \tag{56}$$

where, as usual, $u_w^0=\epsilon/m$. If for γ_{abc} we use the linear approximation (48), we can easily see that (56) indeed corresponds to the discussed contribution to the amplitude provided that $f_{2,0}=1$. Thus, under gravity the value of one more form factor at zero momentum transfer t is fixed. This corresponds to conservation of angular momentum. This was pointed out long ago by Kobzarev and Okun'[35] and Hehl *et al.*[36]

Let us now go back to the convection term in (54). As in electrodynamics, when we go over to spinors in the rest frame, the term of first order in spin can be written as

$$-\frac{p^\mu p^\nu}{8\epsilon}\,\frac{1}{m}\,\frac{u^0}{u^0+1}(\mathbf{s}\cdot\mathbf{v}\times\boldsymbol{\nabla})h_{\mu\nu}. \tag{57}$$

Using (25) and (48), we obtain

$$p^\mu\nabla_k h_{\mu\nu}\to -p^\mu(-\partial_k h_{\mu\nu}+\partial_\mu h_{k\nu})\to -2p^a\gamma_{ak\nu}.$$

Thus, the expression (57) can be written in terms of the Ricci coefficients as follows:

$$\frac{1}{u_w^0}\,\frac{u^0}{u^0+1}\,\epsilon^{mnk}s^m v^n u^a u^c\gamma_{akc}. \tag{58}$$

Clearly, the sum of (56) and (58) reproduces the Lagrangian (47).

7. EQUATIONS OF MOTION OF A SPINNING PARTICLE IN A GRAVITATIONAL FIELD. SECOND ORDER IN THE SPIN

Let us now investigate the effects of second order in the spin in the equations of motion in a gravitational field. In the case of a binary star these effects are of the same order of magnitude as the spin–spin interaction when the spins of the components of the system are comparable.[23] The influence of the latter on the characteristics of gravitational radiation becomes noticeable for a system of two extreme black holes.[20] Correspondingly, second-order spin effects in the equations of motion become substantial if at least one component of a binary is an extreme black hole.[23] Therefore, it is interesting to study these effects not only from the theoretical viewpoint: they can be detected (at least in principle) with the gravitational wave detectors under construction.

An obvious source of second-order spin effects is the term

$$L_{2sg}=-f_{1,2}\frac{1}{8\epsilon}p^\mu p^\nu(Sq)^2h_{\mu\nu} \tag{59}$$

in the amplitude (54). Due to the relation

$$p^\mu p^\nu q_\alpha q_\beta h_{\mu\nu}=p^\mu p^\nu(q_\alpha q_\beta h_{\mu\nu}+q_\mu q_\nu h_{\alpha\beta}-q_\alpha q_\nu h_{\mu\beta} -q_\beta q_\mu h_{\nu\alpha})\to 2p^\mu p^\nu R_{\mu\alpha\nu\beta},$$

the Lagrangian (59) is written in terms of the Riemann tensor:

$$L_{2sg}=-\frac{\kappa}{2\epsilon}u^a S^b u^c S^d R_{abcd}. \tag{60}$$

Instead of $f_{1,2}$, we have introduced a dimensionless parameter κ:

$$f_{1,2}=\frac{\kappa}{2m^2}.$$

At this point it is convenient to use the Petrov representation for the components of the Riemann tensor:[34]

$$E_{kl}=R_{0k0l},\quad E_{kl}=E_{lk},$$
$$C_{kl}=\tfrac{1}{4}\epsilon_{kmn}\epsilon_{lrs}R_{mnrs},\quad C_{kl}=C_{lk}, \tag{61}$$

$$B_{kl}=\frac{1}{2}\epsilon_{lrs}R_{0krs}, \quad B_{kk}=0.$$

We limit our discussion to the case of a sourceless gravitational field. Then, at $R_{ab}=0$, we can simplify (61) still further:

$$C_{kl}=-E_{kl}, \quad B_{kl}=B_{lk}, \quad E_{kk}=C_{kk}=0. \tag{62}$$

Finally, we arrive at the following interaction Lagrangian that is quadratic in the spin:

$$L_{2sg}=-\frac{\kappa}{2\epsilon}\Bigg[(2\mathbf{u}^2+1)E_{kl}-2\left(2-\frac{1}{u_0+1}\right)u_k u_m E_{lm}$$
$$+\delta_{kl}u_m u_n E_{mn}+\frac{1}{(u_0+1)^2}u_k u_l u_m u_n E_{mn}$$
$$-2u_0\epsilon_{kmn}u_m B_{nl}+\frac{2}{u_0+1}u_k u_m \epsilon_{lrn}u_r B_{mn}\Bigg]$$
$$\times\left(s_k s_l-\frac{1}{3}\delta_{kl}\mathbf{s}^2\right). \tag{63}$$

To avoid any misunderstanding, we note that all three-dimensional indices in Eq. (64) (and in Eq. (65)) are in fact contravariant.

As in electrodynamics, along with the bare interaction (63) there is an induced interaction quadratic in the spin. The explicit expression can be obtained most easily by setting $g=2$ in the electromagnetic formula (36) and by introducing the substitution (45). We also take into account the correspondence

$$q_i\gamma_{abc}u^c=(q_i\gamma_{abc}-q_c\gamma_{abi})u^c$$
$$\to i(\partial_i\gamma_{abc}-\partial_c\gamma_{abi})u^c\to i\,R_{abci}u^c.$$

Finally, using (61) and (62), we obtain the following result for the induced interaction:

$$L^i_{2sg}=\frac{1}{2\epsilon}\Bigg\{\left(2\mathbf{u}^2-\frac{u^0-1}{u^0+1}\right)E_{kl}-2\left[2-\frac{1}{u_0+1}-\frac{1}{(u_0+1)^2}\right]$$
$$\times u_k u_m E_{lm}+\left[1-\frac{1}{(u_0+1)^2}\right]\delta_{kl}u_m u_n E_{mn}+\frac{1}{(u_0+1)^2}$$
$$\times u_k u_l u_m u_n E_{mn}-2\left(u_0-\frac{1}{u_0+1}\right)\epsilon_{kmn}u_m B_{nl}$$
$$+\frac{2}{u_0+1}u_k u_m\epsilon_{lrn}u_r B_{mn}\Bigg\}s_k s_l. \tag{64}$$

As in the electromagnetic case, the induced interaction tends to zero in the nonrelativistic limit $\sim v/c$ and the spin factor $s_k s_l$ in it is not an irreducible tensor.

The asymptotic behavior of L_{2sg} and L^i_{2sg} is the same: both Lagrangians increase linearly with energy. However, in this case too the coefficient in the bare interaction can be chosen in such a way, $\kappa=1$, that the total Lagrangian of second order in the spin decreases (as well as the analogous interaction in electrodynamics) when the energy tends to infinity. At $\kappa=1$,

$$L_{2sg}+L^i_{2sg}=-\frac{1}{\epsilon(u^0+1)}\Bigg(u^0E_{kl}-\frac{1}{u_0+1}u_k u_m E_{lm}$$
$$+\frac{1}{2(u_0+1)}\delta_{kl}u_m u_n E_{mn}+\epsilon_{kmn}u_m B_{nl}\Bigg)s_k s_l. \tag{65}$$

8. GRAVIMAGNETIC MOMENT: MULTIPOLES OF BLACK HOLES

There is a profound analogy between the Lagrangian of the interaction of the magnetic moment which is linear in the spin and the electromagnetic field,

$$\mathscr{L}_{em}=-\frac{eg}{4m}F_{ab}S^{ab}, \tag{66}$$

and the bare gravitational Lagrangian (60), which is quadratic in the spin.[11] (Here it proves more convenient to write the gravitational Lagrangian (like $\mathscr{L}_{em}$) for the proper time τ rather than the world time t, i.e., to multiply (60) by ϵ/m.) This analogy is based on the following observation. It is well-known that the canonical momentum $i\partial_\mu$ enters into relativistic wave equations for a particle in electromagnetic and gravitational external fields via the combination

$$\Pi_\mu=i\partial_\mu-eA_\mu-\tfrac{1}{2}\Sigma^{ab}\gamma_{ab\mu}.$$

The structure of the commutator (or Poisson bracket in the classical limit) implies that

$$[\Pi_\mu,\Pi_\nu]=-i(eF_{\mu\nu}-\tfrac{1}{2}\Sigma^{ab}R_{ab\mu\nu})$$

that in a sense $-(1/2)\Sigma^{ab}R_{ab\mu\nu}$ plays the same role in gravity as $eF_{\mu\nu}$ in electromagnetism. It is quite natural then that the gravitational analog of the electromagnetic spin interaction (66) is

$$\mathscr{L}_{gm}=\frac{\kappa}{8m}R_{abcd}S^{ab}S^{cd}. \tag{67}$$

Clearly, (67) and (60) coincide (to within the factor ϵ/m). This become evident if we bear in mind that $S^{ab}=\epsilon^{abcd}S_c u_d$ and that

$$\tilde{R}_{abcd}=\tfrac{1}{4}\,\epsilon^{ef}_{ab}\,\epsilon^{gh}_{cd}\,R_{efgh}-R_{abcd},$$

which is true for a sourceless gravitational field.

By analogy with the magnetic moment

$$\frac{eg}{2m}S^{\mu\nu},$$

it is natural to introduce the gravimagnetic moment

$$-\frac{\kappa}{2m}S^{ab}S^{cd}.$$

The gravimagnetic ratio κ, like the gyromagnetic ratio g in electrodynamics, can generally take any value. However, it is quite natural that under gravity the value $\kappa=1$ is just as special, as $g=2$ is in electrodynamics. In any case, at $g=2$ and $\kappa=1$ the spin equations of motion are the simplest.

For a classical object the values of both parameters, g and κ, generally depend on the various properties of the

object. For black holes the situation is different, however. From an analysis of the Kerr–Newman solution it follows that the gyromagnetic ratio of a charged rotating black hole is universal (and equal to that of the electron): $g=2$ (see Ref. 37).

Let us show that for a Kerr black hole the gravimagnetic ratio is $\kappa=1$. In fact, this value follows from the analysis of the motion of the spin of a black hole in an external field done by Thorne and Hartle[19] (although they did not formulate this statement explicitly). Here we present an independent and, in our opinion, simpler derivation of this important result.

At great distances from a Kerr hole the hole can be considered a point source of a weak gravitational field. To first order in the field of the hole at rest, the Lagrangian density corresponding to the interaction (60) can be written as

$$L=\frac{\kappa}{4m}(\mathbf{s}\cdot\nabla)^2 h_{00}\delta(\mathbf{r}). \tag{68}$$

The correction to the energy-momentum tensor induced in this manner has only one component,

$$\delta T_{00}=-\frac{\kappa}{2m}(\mathbf{s}\cdot\nabla)^2\delta(\mathbf{r}). \tag{69}$$

Let us find the corresponding correction to the 00-component of the metric. In the gauge

$$h^{\mu\nu}{}_{,\nu}=0, \quad \overline{h}_{\mu\nu}=h_{\mu\nu}-\tfrac{1}{2}\eta_{\mu\nu}h^{\alpha}_{\alpha}, \tag{70}$$

the static Einstein equation for h_{00} in the linear approximation is

$$\Delta h_{00}=8\pi k T_{00}.$$

The induced correction to metric is

$$h_{00}=\kappa\frac{k}{m}(\mathbf{s}\cdot\nabla)^2\frac{1}{r}. \tag{71}$$

Let us now compare this corrections with the corresponding contribution to the Kerr metric. In Boyer–Lindqvist coordinates this metric is

$$ds^2=\left(1-\frac{r_g r}{\Sigma}\right)dt^2-\frac{\Sigma}{\Delta}dr^2-\Sigma\,d\theta^2-\left(r^2+a^2+\frac{r_g r a^2}{\Sigma}\sin^2\theta\right)r^2\sin^2\theta+\frac{2r_g r a}{\Sigma}\sin^2\theta d\phi\,dt, \tag{72}$$

where $\Delta=r^2-r_g r+a^2$, $\Sigma=r^2+a^2\cos^2\theta$, and $\mathbf{a}=\mathbf{s}/m$. For $r_g=0$ the metric (72) describes a flat space in spheroidal coordinates.[34] At the same time, in a flat space Cartesian coordinates correspond to the gauge (70). The transition from spheroidal coordinates to Cartesian coordinates is carried out with the required accuracy by the substitution

$$\mathbf{r}\to\mathbf{r}+\frac{\mathbf{a}(\mathbf{a}\cdot\mathbf{r})-\mathbf{r}a^2}{2r^2}.$$

Obviously, in Cartesian coordinates the spin-dependent part of the 00-component of the metric

$$g_{00}=1-\frac{r_g}{r}+\frac{r_g a^2}{2r^3}(3\cos^2\theta-1),$$

coincides with h_{00} of (71) at $\kappa=1$. A somewhat more complicated discussion of the spatial components of the Kerr metric leads to the same result, $\kappa=1$.

Note that the motion of a Kerr black hole in an external gravitational field is not described by the Papapetrou equation even we ignore the problem of a spin–orbit interaction which is linear in the spin. The important thing is that this equation refers to the case $\kappa=0$ (see Ref. 14).

Reasoning along similar lines, we can prove that for a charged Kerr hole the gravimagnetic ratio κ is also unity. Moreover, it can be proved that the electric quadrupole moment of a charged Kerr hole also equals

$$Q=-2\frac{es^2}{m^2},$$

a value at which the interaction that is quadratic in spin decreases with increasing energy. We can also demonstrate that the higher multipoles of a charged Kerr hole have values that guarantee that the interaction of any order in the spin (but, of course, linear in the external field) asymptotically decreases with increasing energy.

We are grateful to I. A. Koop, R. A. Sen'kov, and Yu. M. Shatunov for useful discussions. We are also grateful to R. Ruffini, D. Bini, G. Gemelli, and P. Menotti and to the JETP referee for their remarks due to which, we believe, the present paper has become more comprehensible. The work was supported by the Russian Fund for Fundamental Research (Grant No. 95-02-04436-a.

*)E-mail: pomeransky@vxinpz.inp.nsk.su
†)E-mail: khriplovich@inp.nsk.su

1)The authors of this paper overlooked a misprint in the answer for the induced quadrupole moment. As a result the value presented in Ref. 31 is half the correct value.

[1] L. H. Thomas, Nature **117**, 514 (1926); Philos. Mag. **3**, 1 (1926).
[2] W. de Sitter, Mon. Not. R. Astron. Soc. **77**, 155, 181 (1916).
[3] L. Schiff, Phys. Rev. Lett. **4**, 435 (1959).
[4] J. Frenkel, Z. Phys. **37**, 243 (1926).
[5] V. Bargman, L. Michel, and V. Telegdi, Phys. Rev. Lett. **2**, 435 (1959).
[6] A. Papapetrou, Proc. R. Soc. London, Ser. A **209**, 248 (1951).
[7] A. Barducci, R. Casalbuoni, and L. Lusanna, Nuovo Cimento A **35**, 389 (1976).
[8] F. Ravndal, Phys. Rev. D **21**, 2823 (1980).
[9] P. L. Nash, J. Math. Phys. **25**, 2104 (1984).
[10] U. Heinz, Phys. Lett. B **144**, 228 (1984); Ann. Phys. (San Diego) **161**, 48 (1985).
[11] I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **96**, 385 (1989) [Sov. Phys. JETP **69**, 217 (1989)].
[12] J. W. van Holten, Nucl. Phys. B **356**, 3 (1991).
[13] R. H. Rietdijk and J. W. van Holten, Class. Quantum Grav. **9**, 575 (1992).
[14] K. Yee and M. Bander, Phys. Rev. D **48**, 2797 (1993).
[15] J. P. Costella and B. H. J. McKellar, Int. J. Mod. Phys. A **9**, 461 (1994).
[16] M. Chaichian, R. Gonzales Felipe, and D. Louis Martinez, Preprint hep-th/9601119.
[17] Ya. S. Derbenev and A. M. Kondratenko, Zh. Éksp. Teor. Fiz. **64**, 1918 (1973) [Sov. Phys. JETP **37**, 968 (1973)].
[18] K. Heinemann, Preprint physics/9611001.
[19] K. S. Thorne and J. B. Hartle, Phys. Rev. D **31**, 1815 (1985).

[20] L. E. Kidder, C. M. Will, and A. G. Wiseman, Phys. Rev. D **47**, R4183 (1993).
[21] L. Blanchet, T. Damour, B. R. Iyer, C. M. Will, and A. G. Wiseman, Phys. Rev. Lett. **74**, 3515 (1995).
[22] H. T. Cho, E-print archive gr-qc/9703071.
[23] I. B. Khriplovich and A. A. Pomeransky, Phys. Rev. Lett. **216**, 7 (1996); E-printarchive gr-qc/9602004.
[24] B. M. Barker and R. F. O'Connell, Gen. Relativ. Gravit. **5**, 539 (1974).
[25] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed., Pergamon Press, Oxford (1991).
[26] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Fields*, McGraw-Hill, New York (1965).
[27] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 248 (1951).
[28] W. Pauli, *Les théories quantique du magnétisme. L'électron magnétique. 6éme Conseil de Physique Solvay*, Bruxelles, 1930, p. 175.
[29] T. O. Niinikoski and R. Rosmanith, Nucl. Instrum. Methods A **225**, 460 (1987).
[30] I. B. Khriplovich, *Parity Nonconservation in Atomic Phenomena*, Gordon & Breach, Philadelphia (1991).
[31] I. B. Khriplovich, A. I. Milstein, and R. A. Sen'kov, Phys. Lett. A **221**, 370 (1996); Zh. Éksp. Teor. Fiz. **111**, 1935 (1997) [JETP **84**, 1054 (1997)].
[32] S. Weinberg, in *Lectures on Elementary Particles and Quantum Field Theory*, S. Deser, M. Grisaru, and H. Pendleton (eds.), M.I.T. Press, Cambridge, MA (1970).
[33] S. Ferrara, M. Porrati, and V. L. Telegdi, Phys. Rev. D **46**, 3529 (1992).
[34] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed., Pergamon Press, Oxford (1975).
[35] I. Yu. Kobzarev and L. B. Okun', Zh. Éksp. Teor. Fiz. **43**, 1904 (1962) [Sov. Phys. JETP **16**, 1343 (1962)].
[36] F. W. Hehl, A. Macias, E. W. Mielke, and Yu. N. Obukhov, E-print archive gr-qc/9706009.
[37] B. Carter, Phys. Rev. **174**, 1559 (1968).
[38] Ya. I. Azimov and R. M. Ryndin, in *Proc. 21st Winter Workshop at St. Petersburg Nuclear Physics Institute* [in Russian], St. Petersburg (1992); E-print archive hep-ph/9710433; 9707468.

Translated by Eugene Yankovsky

Mass operator of an axion in a crossed field

V. V. Skobelev

Moscow State Industrial University, 109280 Moscow, Russia
(Submitted 31 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1558–1565 (May 1998)

The dominant one-loop electron contribution to the mass operator of an axion in a crossed field in the asymptotic limits of the parameters q^2/m_e^2 and $\chi=\sqrt{e^2(qF^2q)/m_e^3}$ is calculated. The corresponding electromagnetic mass of the axion is compared with the quantum-chromodynamic mass due to mixing with π^0. Expressions are derived for the probability of pair creation $a\to e^+e^-$, and the fundamental conclusion is reached that refractive effects are present in the propagation of an axion in an external electromagnetic field. © *1998 American Institute of Physics.* [S1063-7761(98)00205-4]

The axion is a hypothetical particle whose existence may explain the absence of violations of CP invariance in strong interactions on the basis of natural dynamical considerations. Specifically, in the Peccei–Quinn scheme[1] the pseudoscalar axion field a is introduced as initially massless with a Lagrangian for the interaction with the quantum chromodynamics (QCD) color field (with a kinetic term) in the form

$$\mathscr{L}_a=\frac{1}{2}\frac{\partial a}{\partial x_\mu}\frac{\partial a}{\partial x^\mu}-\frac{\alpha_s}{8\pi f_a}aG_{b\mu\nu}\tilde{G}_b^{\mu\nu}, \tag{1}$$

where α_s is an effective dimensionless coupling constant in the strong interactions, b is the color index, $\tilde{G}$ is the dual tensor, f_a is a constant with dimensions of energy (the Peccei–Quinn energy scale). The global symmetry U_{PQ} of (1) consists in invariance of the sum of the Lagrangian $\mathscr{L}_a$ and the effective QCD Lagrangian $\mathscr{L}_s$ (Ref. 2) with respect to the corresponding transformation, reducing for the axion field to the shift $a\to a+a_0$.

The main idea of Peccei and Quinn is that in quantum theory the axion acquires a dynamical mass due to mixing with π^0, since the transition axion$\to$2 gluons$\to$2 quarks $\to\pi^0$ meson has nonzero amplitude. The corresponding term in the Lagrangian $(1/2)m_a^2a^2$ violates the U_{PQ} symmetry of (1); however, this means that in general the low-energy Lagrangian (1) contains an effective potential $V(a)$ ($V_{a\to 0}\simeq(1/2)m_a^2a^2$). The parameters of $V(a)$ can be chosen such that the sum $\mathscr{L}_s+\mathscr{L}_a$ does not contain CP-invariant terms (for more details, see the review in Ref. 3). In other words, the axion is a pseudoscalar Goldstone boson, that is allowed upon spontaneous U_{PQ} symmetry breaking in (1), and f_a is the energy scale of this symmetry breaking, where $f_a\gtrsim 10^{10}$ GeV (although arguments exist both in favor of a larger[4] and a smaller[5] value[1)] of the lower limit).

An equivalent scheme for realizing this mechanism consists in introducing a direct axion–fermion coupling[6] of the form

$$\mathscr{L}=\frac{c_f}{2f_a}(\bar{\Psi}\gamma_\alpha\gamma^5\Psi)\frac{\partial a}{\partial x_\alpha}, \tag{2}$$

which is preferable[3] (here c_f is a model-dependent dimensionless constant of order unity). This gives rise to the existence of self-energy diagrams with a fermion loop (see Fig. 1). As the axion propagates in an external electromagnetic field, as a result of its interaction with the charged fermion propagator the axion can acquire an electromagnetic mass δm_a, and it is of interest to compare it with the QCD mass m_a. The latter, according to astrophysical data, can vary over a wide range: 10^{-5} eV$\lesssim m_a\lesssim$10 eV. For reasonable values of the field and energy of the axion (in the same reference frame) it is possible to restrict the discussion to the contribution of the electron loop ($f=e$), which corresponds to the lowest charged fermion mass.

In a more general formulation it is necessary to find the mass operator of the axion in an electromagnetic field, whose imaginary part determines the probability of creation of an e^+e^- pair in this field, which is of independent interest.

In the calculations that follow, we employ the invariant crossed-field technique developed by Ritus.[7] In such a field, the two invariants

$$h=\frac{e^2(F^2)_\alpha^a}{m^4},\quad g=\frac{e^2}{m^4}e_{\mu\nu\alpha\beta}F^{\mu\nu}F^{\alpha\beta}$$

are equal to zero (e and m are the charge and mass of the electron). However, if $\chi^2\gg h,g$, where

$$\chi=\frac{1}{m^3}\sqrt{e^2(qF^2q)}, \tag{3}$$

the method also yields adequate results for arbitrary weakly-varying external electromagnetic fields.

In accordance with the form of the interaction Lagrangian (2), the single-loop electron contribution to the axion mass operator has the form

$$M^2=-\frac{ic_e^2}{4f_a^2}\int d^4z e^{i(qz)}\mathrm{Tr}[\hat{q}\gamma^5G(z)\hat{q}\gamma^5G(-z)], \tag{4}$$

where q is the momentum of the external line, $G(z)$ is the coordinate-difference–dependent part of the total Green's function of an electron in the crossed field

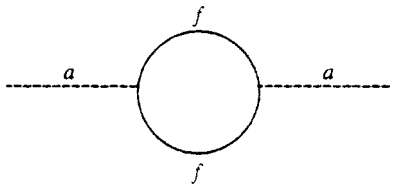

FIG. 1.

$$S(x,y)=\exp\left[ie\int_x^y (dx'A(x'))\right]G(x-y), \tag{5}$$

which in the proper-time representation is equal to ($eF\to F$)

$$G(z)=-\frac{1}{(4\pi)^2}\int_0^\infty \frac{ds}{s^2}\exp\left[-\frac{iz^2}{4s}-ism^2 - is\,\frac{(zF^2z)}{12}\right]\left[m+\frac{\hat z}{2s}-\frac{s}{3}\,(\gamma F^2 z) + \frac{ms}{2}\,(\gamma F\gamma)-\frac{i\gamma^5(\gamma\tilde F z)}{2}\right], \tag{5a}$$

$\tilde F_{\alpha\beta}=(1/2)e_{\alpha\beta\mu\nu}F^{\mu\nu}$ is the dual tensor, and $\gamma^5=-i\gamma^0\gamma^1\gamma^2\gamma^3$ (the phase factor in Eq. (5) in the loop with two vertices cancels out). Taking Eq. (5a) into account, expression (4) can be written in the form

$$M^2=-\frac{ic_e^2}{4(4\pi)^4 f_a^2}\int_0^\infty\frac{ds_1}{s_1^2}\int_0^\infty\frac{ds_2}{s_2^2}\,\mathrm{Tr}\left\{\hat q\gamma^5\left[m+\frac{\hat z}{2s_1} - \frac{s_1}{3}\,(\gamma F^2 z)+\frac{ms_1}{2}\,(\gamma F\gamma)-\frac{i\gamma^5(\gamma\tilde F z)}{2}\right]\hat q\gamma^5 \times\left[m-\frac{\hat z}{2s_2}+\frac{s_2}{3}\,(\gamma F^2 z)+\frac{ms_2}{2}\,(\gamma F\gamma) + \frac{i\gamma^5(\gamma\tilde F z)}{2}\right]\right\}\mathscr{T}\exp[-im^2(s_1+s_2)], \tag{6}$$

where z here is understood to mean the differential operator $z^\alpha=-i\partial/\partial q_\alpha$ acting on the function $\mathscr{T}$ that appears as a result of integration over the coordinate z:

$$\mathscr{T}=16i\pi^2\beta^2\exp\left[i\beta\left(q^2-\frac{1}{3}\,s_1s_2(qF^2q)\right)\right], \qquad \beta=\frac{s_1s_2}{s_1+s_2}. \tag{7}$$

After taking the trace and calculating quadratic combinations like z^2, (zF^2z), $(qF^2z)^2$, etc. with the help of the relation

$$z_\mu z_\nu\to -i(2\beta)\left(g_{\mu\nu}-\frac{1}{3}\,s_1s_2F^2_{\mu\nu}\right)+(2\beta)^2\left[q_\mu-\frac{1}{3}\,s_1s_2(qF^2)_\mu\right]\left[q_\nu-\frac{1}{3}\,s_1s_2(qF^2)_\nu\right]$$

we obtain

$$M^2=\frac{c_e^2}{16\pi^2f_a^2}\int_0^\infty\int_0^\infty\frac{ds_1ds_2}{(s_1+s_2)^2}\exp\left\{-im^2(s_1+s_2) + i\beta\left[q^2-\frac{1}{3}\,s_1s_2(qF^2q)\right]\right\}\left\{2i\beta\left[-\frac{q^2}{2s_1s_2} - \frac{1}{3}\left(\frac{s_1}{s_2}+\frac{s_2}{s_1}-1\right)(qF^2q)\right]-m^2[q^2 - 2s_1s_2(qF^2q)]+(2\beta)^2\left[-\frac{q^4}{4s_1s_2} + \frac{1}{6}\left(\frac{s_1}{s_2}+\frac{s_2}{s_1}+\frac{5}{2}\right)q^2(qF^2q) - \frac{1}{9}\,s_1s_2\left(\frac{s_1}{s_2}+\frac{s_2}{s_1}+\frac{5}{2}\right)(qF^2q)^2\right]\right\}. \tag{8}$$

The given expression diverges at small s_1 and s_2. This is a consequence of the indeterminacy of the single-loop contributions to M^2 in the absence of a field due to multivariance of the axion interactions. The ''field'' contribution

$$M_F^2=M^2-M^2|_{F=0}, \tag{9}$$

which is of interest to us, converges.

Transforming to dimensionless coordinates

$$v=(s_1+s_2)m^2, \quad u=\frac{4s_1s_2}{(s_1+s_2)^2}, \tag{10}$$

after some additional transformations we obtain the final general result

$$M_F^2=\frac{c_e^2m^4}{64\pi^2f_a^2}\int_0^\infty\frac{dv}{v}\int_0^1\frac{du}{(1-u)^{1/2}}\exp\left(-iv+i\tilde q^2\,\frac{uv}{4}\right) \times\left\{\left(1-\exp\left(-\frac{i}{48}\,\chi^2u^2v^3\right)\right)\tilde q^2\left(1+\frac{u}{4}\,\tilde q^2+\frac{i}{v}\right) + \frac{1}{2}\,\chi^2v\,\exp\left(-\frac{i}{48}\,\chi^2u^2v^3\right)\left[-i\left(\frac{4}{3}-u\right) + uv+\frac{1}{3}\,\tilde q^2uv\left(1+\frac{u}{8}\right)-\frac{1}{18}\,\chi^2u^2v^3\left(1+\frac{u}{8}\right)\right]\right\}, \tag{11}$$

where $\tilde q^2=q^2/m^2$ and χ is given by Eq. (3).

Calculation of the integrals in expression (11) is not possible; therefore it is necessary to restrict the discussion to special values of the parameters $\tilde q^2$ and χ.

1. $\chi\ll 1$, $\tilde q^2$ anywhere in the spacelike region $\tilde q^2<0$. The leading contribution here is obtained by formally expanding expression (11) in χ^2. After some simple calculations we find

$$M_F^2=\frac{c_e^2m^4\chi^2}{96\pi^2f_a^2}\,\frac{(\Delta-\tilde q^2)^2}{\Delta^4}\left\{3\left[2+(4-\Delta)\,\frac{\Delta+\tilde q^2}{\Delta-\tilde q^2}\right] + \frac{4}{\Delta}\,\frac{\Delta-\tilde q^2}{\Delta+\tilde q^2}\left[1+\frac{4}{(\Delta-\tilde q^2)^4}\,(\Delta+\tilde q^2) \times(\Delta^2+\tilde q^4)(2\Delta-\tilde q^2)\right]\ln\left(\frac{\Delta+\tilde q^2}{\Delta-\tilde q^2}\right)\right\}, \tag{12}$$

where $\Delta=(\tilde{q}^4-4\tilde{q}^2)^{1/2}$. Analytic continuation into the region $\tilde{q}^2>0$ is realized by familiar methods.[8] This form, however, is uninteresting since it in fact describes virtual effects and in any case makes only a small field correction to the mass operator in the absence of an external field.

2. In the description of the behavior of a physical axion $q^2=m_a^2$ in an external electromagnetic field, by virtue of its negligibly small mass we can in fact set $\tilde{q}^2=0$ in expression (11).

Introducing the complex-valued Hardy–Stokes function

$$f(x)=i\int_0^\infty dt\ \exp\left[-i\left(tx+\frac{t^3}{3}\right)\right], \tag{13}$$

whose imaginary part coincides with the well-known Airy function $\Phi(x)$, after transforming from the variable v to the new variable

$$t=v/x,\quad x=(4/\chi u)^{2/3}, \tag{14}$$

we obtain

$$M_F^2=\frac{c_e^2m^4\chi^2}{384\pi^2f_a^2}\int_0^1\frac{x\,du}{(1-u)^{1/2}}\left[-(4-3u)f+3uxf'+\frac{1}{6}\chi^2u^2\left(1+\frac{u}{8}\right)x^3f'''\right]. \tag{15}$$

Using the differential equation for the function $f(x)$

$$f''-xf=-1, \tag{16}$$

it is more convenient, taking the explicit form of x (14) into account, to rewrite formula (15) in the form

$$M_F^2=\frac{c_e^2m^4\chi^2}{288\pi^2f_a^2}\int_0^1\frac{x\,du}{(1-u)^{1/2}}\times\left[\left(-1+\frac{5}{2}u\right)f+\left(2+\frac{5}{2}u\right)xf'\right]. \tag{17}$$

This expression for arbitrary finite χ has both a real and an imaginary part, where the latter is related to the probability W_F of pair creation by an axion in a crossed field by the unitarity condition

$$\mathrm{Im}\ M_F^2=-q_0W_F. \tag{18}$$

a) $\chi\ll1$. To find the real part it is sufficient to use the expansion

$$f(x)|_{x\gg1}=\frac{1}{x}+\frac{2}{x^4}+\ldots,$$

after which we easily obtain

$$\mathrm{Re}\ M_F^2\simeq-\frac{c_e^2m^4\chi^2}{48\pi^2f_a^2}. \tag{19}$$

This result follows from Eq. (12) for $\tilde{q}^2=0$ and is valid in arbitrary $F=\mathrm{const}$, since the possible difference consists in adding a term of the form $h(m_a^2/m^2)$ to χ^2 in Eq. (19), which is negligibly small.

Next, replacing f by Φ in expression (17) and using the asymptotic limit of the Airy functions as $x\to\infty$,

$$\Phi(x)\simeq\frac{\sqrt{\pi}}{2x^{1/4}}\exp\left(-\frac{2}{3}x^{3/2}\right), \tag{20}$$

employing standard integral formulas[9] we find the following approximate expression for the imaginary part of M_F^2:

$$\mathrm{Im}\ M_F^2\simeq-\left(\frac{3}{2}\right)^{1/2}\frac{c_e^2m^4}{32\pi f_a^2}\chi\exp\left(-\frac{8}{3\chi}\right). \tag{21}$$

Such a dependence (taking Eq. (18) into account) of W_F on χ is of a general character and was noted, for example, in Ref. 7 for the process of pair formation by a photon in a crossed field.

b) $\chi\gg1$. In the treatment of this limit we cannot set $f(x)\to f(0)$ and $f'(x)\to f'(0)$ in (17), since the integral will diverge; therefore the calculations require care. To this end, we split the integral in Eq. (17) into three parts

$$I=I_1+I_2+I_3, \tag{22}$$

$$\begin{Bmatrix}I_1\\I_2\\I_2\end{Bmatrix}=\int_0^1\frac{du\ x}{\sqrt{1-u}}\begin{Bmatrix}\frac{5}{2}uxf'\\2xf'\\\left(-1+\frac{5}{2}u\right)f\end{Bmatrix}. \tag{22a}$$

First note that in the calculation of an integral of the type

$$J=\int_0^1 du\ f(x)\varphi(u) \tag{23}$$

in the asymptotic limit $\chi\to\infty$, taking the argument of (14) into account, the leading term can be obtained by making the substitution $f(x)\to f(0)$ if the remaining part of the integral converges, while the following term of the expansion δJ is formed in the region $u\lesssim\chi^{-1}$, where x is large and $f\simeq1/x$. We then easily find that

$$\delta J\propto1/\chi^{\alpha+1}, \tag{23a}$$

where α is determined by the condition $\varphi|_{u\to0}\propto u^\alpha$. This is done by making the substitution $f\to f'$ in Eq. (23). The asymptotic limit I_1 is then determined in the conventional way by taking the value of the resulting integral,[9] and is equal to

$$I_1=15\sqrt{\pi}f'(0)\frac{\Gamma(2/3)}{\Gamma(1/6)}\left(\frac{4}{\chi}\right)^{4/3}+O\left(\frac{1}{\chi^2}\right). \tag{24}$$

Taking relation (23a) into account we write the term I_2 in the form

$$I_2=2\int_0^1 du\ x^2f'(x)+2\,f'(0)\int_0^1 du\ x^2\left(\frac{1}{\sqrt{1-u}}-1\right)+O\left(\frac{1}{\chi^2}\right).$$

Since

$$f'(x)=-\frac{3}{2}\left(\frac{\chi}{4}\right)^{2/3}u^{5/3}\frac{df}{du},$$

the first integral can be integrated by parts, and the second integral, upon making the substitution $u=1-y^2$, reduces to a tabulated integral[9] giving

$$I_2=-3\left(\frac{4}{\chi}\right)^{2/3}\left[f(0)+\left(\frac{4}{\chi}\right)^{2/3}f'(0)\right]+\int_0^1 du\ xf(x)$$
$$-6f'(0)\left(\frac{4}{\chi}\right)^{4/3}F\left(1,\frac{4}{3};\frac{5}{3};-1\right)+O\left(\frac{1}{\chi^2}\right), \qquad (25)$$

where F is the hypergeometric function.

Taking relation (23a) into account we write the term I_3 in the form

$$I_3=f(0)\int_0^1\frac{du\ x}{\sqrt{1-u}}\left(\frac{5}{2}u-1\right)+3f(0)\left(\frac{4}{\chi}\right)^{2/3}$$
$$-\int_0^1 du\ xf(x)+O\left(\frac{1}{\chi^2}\right). \qquad (26)$$

The first term vanishes identically,[9] and the last two terms cancel when combined with I_2 (25).

Finally, taking expressions (22) and (24)–(26) and the value

$$f'(0)=\frac{\Gamma(2/3)}{2\cdot 3^{1/3}}(1-i\sqrt{3}),$$

into account, we obtain the asymptotic limit of the mass operator (17)

$$M_F^2=\frac{c_e^2m^4}{f_a^2}\frac{6^{2/3}\Gamma(2/3)(1-i\sqrt{3})}{144\pi^2}\left\{5\sqrt{\pi}\frac{\Gamma(2/3)}{\Gamma(1/6)}\right.$$
$$\left.+2F\left(1,\frac{4}{3};\frac{5}{3};-1\right)-1\right\}[\chi^{2/3}+O(1)], \qquad (27)$$

where the value in braces is positive.

Note that in the electrodynamics of a strong field the typical dependence at large χ has the form $\chi^{2/3}$ (e.g., in the asymptotic limit of the probability of the process $\gamma\to e^+e^-$; see Ref. 7) and is universal, similar to the logarithmic dependence at high energies in the absence of a field. As noted in Ref. 7, replacing the logarithmic dependence by a power-law dependence in an external field leads to an increase in the role of radiative corrections at high energies. The situation is analogous in axion interactions since M_F^2 is also proportional to $\chi^{2/3}$ (27).

Turning now to a discussion of the results, let us first walk through a formal comparison of the effective ''mass'' $m_{\rm eff}=|M_F^2|^{1/2}$ with the QCD mass m_a, where in the optimal version we should use expression (27). After some transformations we find to order of magnitude

$$m_{\rm eff}\sim 10^{-6}c_e\chi^{1/3}m_a, \qquad (28)$$

where we have used the relation $m_af_a\simeq 6\times 10^4$ MeV^2. Taking $c_e\sim 1$, we conclude that interaction effects of the axion with the magnetized electron–positron vacuum can dominate at values $\chi\geqslant 10^{18}$, which is unrealizable even in astrophysical situations.

The concrete physical interpretation of the electromagnetic correction δm_a to the axion mass simplifies for reasonable values of $\chi\ll 1$, where $|\mathrm{Re}\ M_F^2|\gg|\mathrm{Im}\ M_F^2|$. It then follows from Eq. (19) that

$$\delta m_a=\sqrt{M_F^2}\simeq i\frac{c_em^2\chi}{4\sqrt{3}\pi f_a}, \qquad (29)$$

i.e., δm_a is purely imaginary. Physically, this implies the presence of refractive effects in the propagation of an axion in an electromagnetic field; however, they have no practical significance in any case due to their obvious smallness. Nevertheless, the given circumstance is fundamental since coupling of the form (2), i.e., pseudovector⊗pseudovector, leads to the absence of refraction for a finite density of the electron gas[3] (in contrast to the coupling pseudoscalar⊗pseudoscalar). As can be seen, in the propagation of an axion in an electromagnetic field allowing for polarization of the electron–positron vacuum, the situation is substantially different.

Equations (18), (21), and (27) also determine the probabilities of e^+e^- pair creation by an axion in an electromagnetic field in the asymptotic limits in the parameter χ, which complements the results of Ref. 4, where, in particular, the probability in a superstrong magnetic field was found, and Ref. 5, which in a number of other questions discussed the annihilation channel in a crossed field.

Overall the results of this work are an indication of the unchanged applicability of existing ideas about the magnitude of the axion mass, even in the presence of external electromagnetic fields.

[1)]There is a typographical error in Ref. 5: relation (26) should read $f\geqslant 10^7$ GeV.

[1] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
[2] C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Lett. B **63**, 334 (1976); R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 177 (1976).
[3] G. G. Raffelt, Phys. Rep. **198**, 1 (1990).
[4] V. V. Skobolev, Yad. Fiz. **60**, 484 (1997) [Phys. At. Nucl. **60**, 410 (1997)].
[5] V. V. Skobolev, Zh. Éksp. Teor. Fiz. **112**, 25 (1997) [JETP **85**, 13 (1997)].
[6] J. E. Kim, Phys. Rev. Lett. **43**, 103 (1979); M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. B **166**, 493 (1980).
[7] V. I. Ritus, Tr. FIAN **111**, 5 (1979).
[8] E. M. Lifshitz and L. P. Pitaevskiĭ, *Relativistic Quantum Theory*, Part 2, Pergamon Press, Oxford (1974).
[9] I. S. Gradshtein and I. M. Ryzhik (eds.), *Tables of Integrals, Series, and Products*, Academic Press, New York (1965).

Translated by Paul F. Schippnick

Anomaly-free quantization of a string in two-dimensional space–time

S. N. Vergeles*)

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Russia
(Submitted 11 December 1997)
Zh. Éksp. Teor. Fiz. **113**, 1566–1578 (May 1998)

An anomaly-free quantum theory of a relativistic string is constructed in two-dimensional space–time. The states of the string are found to be similar to the states of a massless chiral quantum particle. This result is obtained by generalizing the concept of an ''operator'' in quantum field theory. © *1998 American Institute of Physics.* [S1063-7761(98)00305-9]

1. INTRODUCTION

It has recently been asserted in a number of works (see, for example, Refs. 1 and 2) that anomaly-free quantization of some models of two-dimensional gravity is possible. Specifically, Ref. 1 examined a model of two-dimensional gravity[3] which in certain variables was described by the same constraints of the first class as those describing a relativistic bosonic string in two-dimensional space–time:

$$\mathscr{E}=-\mathscr{E}_0+\mathscr{E}_1\approx 0,$$

$$\mathscr{E}_0=\frac{1}{2}\,((\pi_0)^2+(r^{0\prime})^2),\quad \mathscr{E}_1=\frac{1}{2}\,((\pi_1)^2+(r^{1\prime})^2), \tag{1.1a}$$

$$\mathscr{P}=r^{a\prime}\pi_a=r^{0\prime}\pi_0+r^{1\prime}\pi_1\approx 0. \tag{1.1b}$$

Dimensionless quantities are employed. Here $r^a(x)$ and $\pi_a(x)$, $a=0,1$, are canonically conjugate fields on a one-dimensional manifold, so that the nonzero commutation relations are

$$[r^a(x),\pi_b(y)]=i\delta^a_b\delta(x-y). \tag{1.2}$$

A prime or overdot signifies a derivative $\partial/\partial x$ or $\partial/\partial t$, respectively.

The ground state of the theory is determined at this stage of quantization. This makes it possible to perform normal ordering of the operator in the constraints (1.1). The determined normal ordering in the constraints can lead in turn to anomalies in the commutators of the constraints. These anomalies partially violate the weak equalities (1.1). To determine the ground state of the field, r^a and π_a are expanded in the modes that arise when solving the Heisenberg equations

$$i\dot{r}^a=[r^a,\mathscr{H}],\quad i\dot{\pi}_a=[\pi_a,\mathscr{H}],$$

$$\mathscr{H}=\int dx\mathscr{E}. \tag{1.3}$$

The solutions of Eqs. (1.2) and (1.3) can be written in the form

$$r^a(t,x)=\int\frac{dk}{2\pi}\,\frac{1}{\sqrt{2|k|}}\,\{c^a_k\exp(-i(|k|t-kx))+c^{a+}_k\exp(i(|k|t-kx))\},$$

$$\pi^a(t,x)=-i\int\frac{dk}{2\pi}\,\sqrt{\frac{|k|}{2}}\,\{c^a_k\exp(-i(|k|t-kx))-c^{a+}_k\exp(i(|k|t-kx))\},$$

$$[c^a_k,c^{b+}_p]=2\pi\eta^{ab}\delta(k-p),\quad [c^a_k,c^b_p]=0. \tag{1.4}$$

Here η^{ab} (below—$\eta^{\mu\nu}$)$=\mathrm{diag}(-1,1)$. We also have the commutation relations

$$[\mathscr{H},c^a_k]=-|k|c^a_k,\quad [\mathscr{H},c^{a+}_k]=|k|c^{a+}_k. \tag{1.5}$$

In conventional quantization the operators c^a_k are annihilation operators, while their hermitian conjugate operators c^{a+}_k are creation operators. The ground state $|0\rangle$ satisfies the conditions

$$c^a_k|0\rangle=0. \tag{1.6}$$

Normal ordering of the operators $(c^{a+}_k,\ c^a_k)$ in the quantities (1.1) means that the creation operators stand to the left of all annihilation operators.

Let us consider the state

$$|k,\ a\rangle=c^{a+}_k|0\rangle. \tag{1.7}$$

It follows immediately from the commutation relations (1.5) that

$$\mathscr{H}|k,\ a\rangle=(|k|+E_0)|k,a\rangle, \tag{1.8}$$

where E_0 is the value of the operator $\mathscr{H}$ for the ground state. The relation (1.8) signifies that the operator $\mathscr{H}$ is positive-definite.

In consequence of Eqs. (1.4) and (1.6) we have for the scalar product of the vectors (1.7)

$$\langle k,\ a|p,\ b\rangle=2\pi\eta^{ab}\delta(k-p). \tag{1.9}$$

Hence it is seen that the metric in the full state space is indefinite.

Next, let us calculate the commutator $[\mathscr{E},\mathscr{P}]$. According to Eq. (1.1) it can be represented as a sum of two terms

$$[\mathscr{E}(x),\ \mathscr{P}(y)]=-[\mathscr{E}_0(x),\ r^{0\prime}\pi_0(y)]+[\mathscr{E}_1(x),\ r^{1\prime}\pi_1(x)]. \tag{1.10}$$

In consequence of Eq. (1.2), both commutators on the right-hand side of Eq. (1.10) are identical to within a change of the index a. These commutators are proportional (to within the

ordering) to the quantities $\mathscr{E}_0$ and $\mathscr{E}_1$, respectively. As is well known, normal ordering of the operators in these commutators leads to anomalies.

Indeed, it follows from the commutation relations (1.4) that the correspondences $c_k^0 \leftrightarrow c_k^{1+}$ and $c_k^{0+} \leftrightarrow c_k^1$ establish an isomorphism of the Heisenberg algebras H_0 and H_1, whose generators are $(c_k^0,\ c_k^{0+})$ and $(c_k^{1+},\ c_k^1)$, respectively. In this case the normal ordering of the operators in the algebra H_1 is transformed by the indicated isomorphism into antinormal ordering in the algebra H_0. It is known that in problems of the type we are studying normal and antinormal orderings lead to anomalies that differ only in sign. Therefore the contribution of the first commutator on the right-hand side of the Eq. (1.10) to the anomaly will be $-A$ and that of the second will be A. But, since a minus sign stands in front of the first commutator in Eq. (1.10), the anomaly in Eq. (1.10) equals $-(-A)+A=2A$.

Let us now examine the problem from a different point of view.

In Ref. 1 it is asserted that in the present theory the positive-definiteness condition (1.8) for the operator $\mathscr{H}$ is not necessary. The initial requirement of the theory is satisfaction of the weak equalities (1.1). Therefore we have the right to reject the quantization conditions (1.6) and replace them with the conditions

$$c_k^{0+}|0\rangle=0, \quad c_k^1|0\rangle=0. \tag{1.11}$$

Under the quantization conditions (1.11) the basis of the full Fock space of the theory consists of vectors of the form

$$c_{k_1}^0 \ldots c_{k_m}^0 c_{p_1}^{1+} \ldots c_{p_n}^{1+}|0\rangle. \tag{1.12}$$

It follows from the commutation relations (1.4) that the scalar product of the states (1.12) is positive-definite. Moreover, there is no anomaly in the operator algebra (1.1).

Indeed, under the conditions (1.11) normal ordering consists of arranging the operators $(c_k^0,\ c_k^{1+})$ to the left of all operators $(c_k^{0+},\ c_k^1)$. This means that normal ordering occurs in both Heisenberg algebras H_0 and H_1. Now, with normal ordering both commutators in (1.10) make the same contribution, equal to A, to the anomaly. If we take account of the minus sign in front of the first commutator on the right-hand side of the equality (1.10), the total anomaly is $-A+A=0$.

The absence of an anomaly in the operator algebra $(\mathscr{E},\mathscr{P})$ makes it possible to satisfy all weak equalities $\mathscr{E}\approx 0$ and $\mathscr{P}\approx 0$. Two physical states which the operators $\mathscr{E}$ and $\mathscr{P}$ annihilate are presented in Ref. 1:

$$\Psi_{\text{gravity}}(r^a)=\exp\pm\frac{i}{2}\int dx\,\varepsilon_{ab}r^a r^{b\prime}.$$

In the present paper we shall likewise reexamine the quantization conditions for a relativistic string in two-dimensional space–time. In so doing, we shall determine the space of physical states with a positive-definite scalar product. The nonphysical states are not studied in the theory. The physical states annihilate all constraints of the first class, i.e., all operators are Virasoro. The physical states are characterized by a continuous parameter, which has the meaning of momentum. However, in our theory not all dynamical variables are linear operators in the space of physical states. In the proposed theory the states of a relativistic string in two-dimensional space–time are found to be identical to the states of a massless chiral particle.

2. RELATIVISTIC BOSONIC STRING IN TWO-DIMENSIONAL SPACE–TIME

Let X^μ, $\mu=0,1$, be coordinates in two-dimensional Minkowski space. Let us examine the Nambu action for a bosonic string

$$S=-\frac{1}{l^2}\int \sqrt{-g}\,d^2\xi=\int d\tau\,\mathscr{L}. \tag{2.1}$$

Here $\xi^a=(\tau,\phi)$ are the parameters of the world sheet of the string and

$$g=\det g_{ab}, \quad g_{ab}=\eta_{\mu\nu}\frac{\partial X^\mu}{\partial\xi^a}\frac{\partial X^\nu}{\partial\xi^b}.$$

Here τ is timelike and ϕ is spacelike. The partial derivatives $\partial/\partial\tau$ and $\partial/\partial\phi$ will be denoted below by an overdot and a prime, respectively. It is easy to show that the generalized momenta $\pi_\mu=\partial\mathscr{L}/\partial\dot{X}^\mu$ satisfy the conditions

$$\mathscr{E}=\frac{l^2}{2}\,\pi_\mu\pi^\mu+\frac{1}{2l^2}\,X^{\mu\prime}X'_\mu\approx 0,$$

$$\mathscr{P}=X^{\mu\prime}\pi_\mu\approx 0. \tag{2.2}$$

The quantities $\mathscr{E}(\phi)$ and $\mathscr{P}(\phi)$ exhaust all constraints of the first class. The Hamiltonian of the system is

$$\mathscr{H}=\int d\phi\,\pi_\mu\dot{\phi}^\mu-\mathscr{L}\approx 0.$$

For this reason, following Dirac, we must employ a generalized Hamiltonian which is an arbitrary linear combination of constraints of the first kind of (2.2)

$$\mathscr{H}_T=\int d\phi(v\mathscr{P}+w\mathscr{E}). \tag{2.3}$$

The equations of motion can be obtained from the variational principle

$$\delta S=\delta\left\{\int d\tau\left(\int d\phi\,\pi_\mu\dot{X}^\mu-\mathscr{H}_T\right)\right\}=0. \tag{2.4}$$

In the case of an open string, when ϕ varies from 0 to π, the variational principle (2.4) gives, besides the equations of motion, the boundary conditions

$$\left(v\,\pi_\mu+w\,\frac{1}{l^2}\,X'_\mu\right)\Big|_{\phi=0,\pi}=0, \tag{2.5}$$

which ordinarily are replaced by the conditions

$$v|_{\phi=0,\pi}=0, \quad X'_\mu|_{\phi=0,\pi}=0. \tag{2.6}$$

For a closed string, besides a boundary condition, there is also a periodicity condition.

Let us study an open string next.

The first step in the quantization process is to postulate the commutation relations for the generalized coordinates and momenta:

$$[X^\mu(\phi),\pi^\nu(\phi')]=i\eta^{\mu\nu}\delta(\phi-\phi'). \tag{2.7}$$

The commutation relations (2.7) and the boundary conditions (2.6) are satisfied if

$$X^\mu(\phi)=\frac{1}{\sqrt{\pi}}\left(x^\mu+i\sum_{n\neq 0}\frac{1}{n}\alpha_n^\mu\cos n\phi\right),$$

$$\pi^\mu(\phi)=\frac{1}{\sqrt{\pi}l}\sum_n\alpha_n^\mu\cos n\phi, \tag{2.8}$$

and the elements $(x^\mu,\ \alpha_n^\mu)$ satisfy the commutation relations

$$[x^\mu,\alpha_n^\nu]=i\eta^{\mu\nu}\delta_n,\quad [x^\mu,x^\nu]=0,$$

$$[\alpha_m^\mu,\alpha_n^\nu]=m\eta^{\mu\nu}\delta_{m+n}. \tag{2.9}$$

Since the quantities (2.8) are real,

$$x^{\mu+}=x^\mu,\quad \alpha_n^{\mu+}=\alpha_{-n}^\mu. \tag{2.10}$$

The constraints (2.2) can be represented as

$$(\mathscr{E}\pm\mathscr{P})(\phi)=\frac{1}{2}(\xi_\pm^\mu(\phi))^2, \tag{2.11}$$

where

$$\xi_\pm^\mu(\phi)=\frac{1}{\sqrt{\pi}}\sum_n\alpha_n^\mu\exp(\mp in\phi). \tag{2.12}$$

Hence it is seen that $\mathscr{E}-\mathscr{P}$ differs from $\mathscr{E}+\mathscr{P}$ by the replacement of ϕ by $-\phi$. This simplifies the analysis, since on the interval $-\pi\leqslant\phi\leqslant\pi$ the quantity $\mathscr{E}+\mathscr{P}$ contains all information about the quantities $\mathscr{E}\pm\mathscr{P}$ on the interval $0\leqslant\phi\leqslant\pi$. Therefore, the Fourier components

$$L_n=\frac{1}{2}\int_{-\pi}^{\pi}d\phi(\mathscr{E}+\mathscr{P})\exp(in\phi) \tag{2.13}$$

are equivalent to the set of quantities (2.2) for $0\leqslant\phi\leqslant\pi$. According to Eqs. (2.11)–(2.13), we have

$$L_n=\frac{1}{2}:\sum_m\alpha_{n-m}^\mu\alpha_{\mu m}:. \tag{2.14}$$

The meaning of the ordering operation in Eq. (2.14) is determined by the quantization method.

Let us also write out expressions for the momentum and angular momentum of a string:

$$P^\mu=\int_0^\pi d\phi\pi^\mu,\quad J^{\mu\nu}=\int_0^\pi d\phi(X^\mu\pi^\nu-X^\nu\pi^\mu). \tag{2.15}$$

With the aid of Eqs. (2.6) and (2.7) we immediately verify that

$$[P^\mu,\ \mathscr{H}_T]=0,\quad [J^{\mu\nu},\ \mathscr{H}_T]=0.$$

This means that the momentum and angular momentum of the string are conserved.

In the currently employed quantization the ground state $|0\rangle$ satisfies the conditions

$$\alpha_m^\mu|0\rangle=0,\quad m\geqslant 0. \tag{2.16}$$

The complete space of states is a linear span of vectors of the form

$$\alpha_{m_1}^{\mu_1}\ldots\alpha_{m_s}^{\mu_s}|0\rangle,\quad m_i<0. \tag{2.17}$$

Therefore all α_m^μ are linear operators in the full space of states. From Eqs. (2.9) and (2.16) it follows that the metric in the state space (2.17) is indefinite. The ordering in Eq. (2.14) means that the operators α_m^μ with $m<0$ are arranged to the left of all operators α_n^μ for $n\geqslant 0$. With this ordering the commutators of Virasoro operators contain anomalies

$$[L_n,L_m]=(n-m)L_{n+m}+\frac{1}{12}D(n^3-n). \tag{2.18}$$

Here D is the dimension of the x space, which in our case is 2. Therefore the most that can be achieved is annihilation of the operators L_n with $n\geqslant 0$. As a result the theory is consistent only for $D=26$. A detailed study of the problems arising with the quantization (2.16) can be found in Ref. 4.

We shall now present the path proposed here for quantization of a two-dimensional string that leads to a self-consistent theory of a string in a space of two dimensions. Our method of quantization of a string is similar to Dirac's method of quantization of the electromagnetic field (see Ref. 5, and also Appendix).

Let

$$x_\pm=x^0\pm x^1,\quad \alpha_m^{(\pm)}=\alpha_m^0\pm\alpha_m^1. \tag{2.19}$$

From Eq. (2.9) we obtain

$$[\alpha_m^{(+)},\alpha_n^{(+)}]=[\alpha_m^{(-)},\alpha_n^{(-)}]=0,$$

$$[\alpha_m^{(+)},\alpha_n^{(-)}]=-2m\delta_{m+n},$$

$$[x_+,x_-]=0,\quad [x_+,\alpha_n^{(+)}]=[x_-,\alpha_n^{(-)}]=0,$$

$$[x_+,\alpha_n^{(-)}]=-2i\delta_n,\quad [x_-,\alpha_n^{(+)}]=-2i\delta_n. \tag{2.20}$$

Let us write the Virasoro operators in the variables $\alpha^{(\pm)}$:

$$L_n=-\frac{1}{2}:\sum_m\alpha_{n-m}^{(+)}\alpha_m^{(-)}:. \tag{2.21}$$

By definition, the ordering operation in Eq. (2.21) signifies that either the elements $\alpha^{(+)}$ are arranged to the left of all elements $\alpha^{(-)}$ or the elements $\alpha^{(-)}$ are arranged to the left of all elements $\alpha^{(+)}$. Both orders are equivalent. Indeed,

$$\sum_m\alpha_m^{(-)}\alpha_m^{(+)}=\sum_m\alpha_m^{(+)}\alpha_m^{(-)}+2\sum_m m.$$

It can be assumed that the last sum is zero, since it can be written as $\zeta(-1)-\zeta(-1)$, where $\zeta(s)$ is the Reimann zeta function. It is known that the zeta function

$$\zeta(s)\equiv\sum_{n=1}^{\infty}n^{-s}$$

possesses a unique analytical continuation to the point $s=-1$ and $\zeta(-1)=-1/12$.

For definiteness, let us choose the same ordering as in Eq. (2.21).

According to Eq. (2.20), we have

$$[L_m,\alpha_n^{(-)}]=-n\alpha_{m+n}^{(-)}. \tag{2.22}$$

One can see from Eqs. (2.20) and (2.22) that the weak inequalities $\alpha_n^{(-)}\approx 0$ and $L_n\approx 0$ are algebraically compatible. For this reason, we determine the physical states as the states satisfying the conditions

$$\alpha_n^{(-)}|\rangle=0, \quad n=0,\ \pm 1,\ ... \tag{2.23}$$

It follows immediately from Eqs. (2.23) and (2.21) that

$$L_n|\rangle=0, \quad n=0,\ \pm 1,\ ... \ . \tag{2.24}$$

for any physical states. The inequalities (2.24) signify that for the quantization (2.23) the Virasoro algebra has no anomalies:

$$[L_n,L_m]=(n-m)L_{n+m}. \tag{2.25}$$

The last formula can also be easily obtained by direct calculation of the commutators, provided that the ordering is assumed to be the same as in Eq. (2.21). The quantization conditions (2.23) are precisely analogous to the quantization conditions (A8) used by Dirac to quantize the electromagnetic field.[5]

We call attention to the fact that states of the form

$$\alpha_n^{(+)}|\rangle, \quad n\neq 0, \tag{2.26}$$

are not studied in this theory, since these states do not satisfy the conditions (2.23). For this reason, the matrix elements of the quantities $\alpha_n^{(+)}$ with $n\neq 0$ with respect to the physical states (2.26) cannot be calculated. Therefore the quantities $\alpha_n^{(+)}$ with $n\neq 0$ cannot be operators in the space of physical states. Hence it follows that observables cannot depend on the elements $\alpha_n^{(+)}$ with $n\neq 0$. In other words, observables must commute with all operators $\alpha_n^{(-)}$. According to Eq. (2.20), there are two such quantities: x_- and $p_+(p_\pm\equiv\alpha_0^{(\pm)})$. Both are real.

Thus we can see that the quantities α_n^μ with $n\neq 0$ are not, generally speaking, linear operators in state space in the conventional sense. Here we adhere to the concept formulated and applied by Dirac in Ref. 5. According to this concept, in quantum field theory linear operators operating in certain linear spaces are replaced by so-called q numbers, which form an associative noncommutative algebra with an involution over the complex numbers. Here we shall formulate Dirac's concept using the conventional mathematical terminology.

Let $\mathscr{A}$ be an associative noncommutative involutory algebra with an identity over the complex numbers. Associativity means that for any elements u, v, and w of the algebra $\mathscr{A}$ and any number c the inequalities

$$(uv)w=u(vw), \quad (cu)v=u(cv)=c(uv).$$

hold. The involution property of the algebra means that there exists a mapping $u\mapsto u^+$ from $\mathscr{A}$ into $\mathscr{A}$ such that

$$(u^+)^+=u, \quad (c_1u+c_2v)^+=\overline{c}_1u^++\overline{c}_2v^+, \quad (uv)^+=v^+u^+$$

for any $u,v\in\mathscr{A}$ and any numbers c_1 and c_2. An overbar signifies complex conjugation. It $u^+=u$ holds, the element u is said to be hermitian.

It is also assumed that the algebra $\mathscr{A}$ has a system of generators $\{\alpha_p\}$ for which all relations are limited by the form of the commutators

$$[\alpha_p,\alpha_{p'}]=c_{pp'}.$$

Here $c_{pp'}$ are complex c numbers (in the Dirac sense).

The definition of involutory algebras (or algebras with involution) and other mathematical definitions presented here can be found in Refs. 6 and 7.

Let V be a vector space with elements $|\Lambda\rangle, |\Sigma\rangle, \ldots$ over the complex numbers and let V^+ be the complex conjugate space, whose elements are denoted by $\langle \ldots |$. There is a one-to-one correspondence between the elements of the spaces V and V^+ such that $c|\Lambda\rangle\leftrightarrow\langle\Lambda|\overline{c}$.

For any two vectors $|\Lambda\rangle$ and $|\Sigma\rangle$ there exist two complex self-conjugate c-number quantities $\langle\Lambda|\Sigma\rangle$ and $\langle\Sigma|\Lambda\rangle$. It is assumed that in the space V there exists a basis $\{|\Lambda\rangle\}$ such that

$$\langle\Lambda|\Sigma\rangle=\delta_{\Lambda\Sigma}. \tag{2.27}$$

If the indices Λ and Σ run through a continuous set, then in Eq. (2.27) $\delta_{\Lambda\Sigma}$ must be interpreted as a delta function. The space V is the space of physical states of the theory.

Let $\mathscr{B}\subset\mathscr{A}$ be a noncommutative involution subalgebra with the identity element. The elements of the subalgebra $\mathscr{B}$ are linear operators in the spaces V and V^+ and, as usual,

$$(u|\Lambda\rangle)^+=\langle\Lambda|u^+, \quad u\in\mathscr{B}.$$

The observables correspond to certain hermitian elements from $\mathscr{B}$. If $u\in\mathscr{A}$ and $u\notin\mathscr{B}$, then the action of the element u on vectors from the spaces V and V^+, generally speaking, is not defined. This distinguishes the Dirac theory from the conventional quantum field theory.

In the theories under study all vectors of the space V are, ordinarily, annihilated by a series of operators of the subalgebra $\mathscr{B}$. Therefore the conditions

$$u_N|\rangle=0, \quad u_{N'}|\rangle=0, ..., |\rangle\in V. \tag{2.28}$$

hold. The indices N, N', ... in Eq. (2.28) run through a certain set J of indices. The conditions (2.28) must be algebraically compatible, i.e., the relations

$$[u_N,u_{N'}]=\sum_{N''}\kappa_{NN',N''}u_{N''},$$

where N, N', $N''\in J$ and $k_{NN',\,N''}$ can be any elements of the algebra $\mathscr{A}$, must hold. Evidently, the operators u_N in Eq. (2.28) do not include the identity or simply unity. We denote by $\mathscr{N}\subset\mathscr{B}$ the subalgebra without identity with the generators $\{u_N\}$, where $N\in J$. Thus $\mathscr{N}$ annihilates the space of physical states V.

Let us now examine the set of elements of the form uv, where $u\in\mathscr{A}$ and $v\in\mathscr{N}$. We denote this set as $\mathscr{N}'$. It is evident from the definition that $\mathscr{N}'$ is a left $\mathscr{A}$-module and a subalgebra in $\mathscr{A}$, but $\mathscr{N}'$ is not a subalgebra in $\mathscr{B}$. Nonetheless, the action of the subalgebra $\mathscr{N}'$ on the space V is defined since it is trivial: $\mathscr{N}'$ annihilates the space V. We note that the commutator $[\mathscr{N}',\mathscr{N}']$ is contained in $\mathscr{N}'$. Indeed, if $r,s\in\mathscr{A}$ and $u,v\in\mathscr{N}$, then

$$[ru,sv]=\{[ru,s]v+s[r,v]u+sr[u,v]\}\in\mathscr{N}',$$

since $[u,v]\in\mathscr{N}$. For this reason, the conditions $\mathscr{N}'V=0$ are algebraically compatible.

Concrete theories can also contain other elements of the algebra $\mathscr{A}$, which are not contained in either $\mathscr{B}$ or $\mathscr{N}'$ and are linear operators on the space V.

A distinguishing feature of the Dirac theory is that it does not treat nonphysical state vectors that do not satisfy the conditions (2.28). Moreover, in the Dirac theory there arises an indefinite metric in the state space. This circumstance can radically alter the theory.

Let us return to the discussion of string theory. In the theory proposed here for a two-dimensional string the algebra $\mathscr{A}$ has generators $\{x_\pm,\alpha_m^{(\pm)}\}$, while the subalgebras $\mathscr{B}$ and $\mathscr{N}$ have generators $\{x_-,p_+,\alpha_m^{(-)}\}$ and $\{\alpha_m^{(-)}\}$, respectively. The Virasoro operators L_n are contained in the subalgebra $\mathscr{N}'$. Note that the algebra of operators L_n is an involutory subalgebra in $\mathscr{N}'$, and since $L_n^+=L_{-n}$ holds, the action of the operators L_n is defined in both spaces V and V^+.

From the definitions (2.15) we obtain the following formulas:

$$(\exp(i\omega J^{01}))\alpha_m^{(\pm)}(\exp(-i\omega J^{01}))=(\exp(\pm\omega))\alpha_m^{(\pm)},$$

$$(\exp(i\omega J^{01}))x_\pm(\exp(-i\omega J^{01}))=(\exp(\pm\omega))x_\pm \quad (2.29)$$

and

$$(\exp(ia_\mu P^\mu))x_\pm(\exp(-ia_\mu P^\mu))=x_\pm+\frac{\sqrt{\pi}}{l}a_\pm,$$

$$(\exp(ia_\mu P^\mu))\alpha_m^{(\pm)}(\exp(-ia_\mu P^\mu))=\alpha_m^{(\pm)}. \quad (2.30)$$

Here ω and a^μ are arbitrary real numbers. It is evident from Eqs. (2.29) and (2.30) that translations and Lorentz transformations preserve the condition (2.23).

Both observables x_- and $p_+=\alpha_0^{(+)}$ are real, and $[x_-,p_+]=-2i$. For this reason, we assume that the physical states are eigenstates of the operator p_+:

$$p_+|k\rangle=2k|k\rangle. \quad (2.31)$$

Here k is a continuous real parameter. According to Eq. (2.29)

$$p_+(\exp(-i\omega J^{01}))=(\exp\,\omega)(\exp(-i\omega J^{01}))p_+. \quad (2.32)$$

Let us formally operate with the relation (2.32) on the state $|k\rangle$. As a result of Eq. (2.31) we obtain

$$p_+(\exp(-i\omega J^{01}))|k\rangle=2ke^{\omega}(\exp(-i\omega J^{01}))|k\rangle. \quad (2.33)$$

The last identity makes it possible to determine the action of the operators $(\exp -i\omega J^{01})$ on the physical states, as follows:

$$(\exp(-i\omega J^{01}))|k\rangle=f_\omega|(\exp\,\omega)k\rangle. \quad (2.34)$$

Here f_ω is a complex number different from zero. If the scalar product on physical state vectors is defined in a Lorentz-invariant manner as

$$\langle k|k'\rangle=k\delta(k-k'),$$

then $|f_\omega|=1$. From Eq. (2.34) it is evident that one can assume

$$k>0. \quad (2.35)$$

The hermitian angular momentum operator can be represented in the form

$$J^{01}=\frac{1}{2}(x_-p_+-x_+p_-)+\frac{i}{4}\sum_{n\neq 0}\frac{1}{n}(\alpha_n^{(-)}\alpha_{-n}^{(+)}-\alpha_n^{(+)}\alpha_{-n}^{(-)}). \quad (2.36)$$

We can see that although the expression (2.36) does not belong to either the subalgebra $\mathscr{B}$ or the subalgebra $\mathscr{N}'$, the action of the quantities $(\exp i\omega J^{01})$ on the space of physical states is nonetheless correctly determined.

According to Eqs. (2.8) and (2.15)

$$P^\mu=\frac{\sqrt{\pi}}{l}\alpha_0^\mu=\frac{\sqrt{\pi}}{2l}\{(\delta_0^\mu+\delta_1^\mu)p_++(\delta_0^\mu-\delta_1^\mu)p_-\}.$$

Therefore, we obtain from Eqs. (2.23) and (2.31)

$$P^\mu|k\rangle=\frac{\sqrt{\pi}}{l}k^\mu|k\rangle,\quad k^\mu=(k,k). \quad (2.37)$$

Thus, as a result of the procedure described above for quantizing a two-dimensional string there arises a system similar to a massless chiral quantum particle in space–time with two dimensions.

3. CONCLUSIONS

Let us note the differences of the main properties of string theory quantized in the conventional manner from those of the string theory proposed in the present paper. In the conventional quantization there exists a state which is invariant under Lorentz transformations. This state is the ground state. In this respect the conventional string theory is similar to the standard quantum field theory of point objects. In such field theories the ground state ordinarily is Lorentz-invariant. Conversely, in our approach there does not exist a state that is invariant under Lorentz transformations. For this reason, the quantum-string theory proposed above is analogous to a quantum theory of a single relativistic particle. Once again there does not exist a Lorentz-invariant quantum state of a single relativistic particle. In order for a Lorentz-invariant state to exist in our theory we would have to introduce a string field and second-quantize the string field. In such a theory the ground state would be Lorentz-invariant, since there would be no real strings in the ground state.

In closing, we note that the quantization method proposed here can be applied to a D-dimensional string. This assertion is based on the fact that in string theory there exists an infinite set of so-called *DDF* operators[4] which commute with all Virasoro operators. The *DDF* operators describe almost all (with the exception of the total momentum of the string) physical degrees of freedom of the string. The independence of Virasoro operators from *DDF* operators signifies that Virasoro operators can be put into the form (2.21), after which the quantization scheme which we have proposed above can be applied. However, the theory is much

more complicated in the D-dimensional case because there exists an infinite set of physical degrees of freedom, contained in the *DDF* operators.

This work was supported by the Russian Fund for Fundamental Research under Grant No. 96-02-17331-a and by the Higher School of Science under Grant No. 96-1596821.

APPENDIX

We shall describe the quantization of a free electromagnetic field, as proposed by Dirac in Ref. 5, following Dirac's concept, which is formulated in Sec. 2. The quantization which we propose for a two-dimensional string is performed in accordance with Dirac's procedure.

The quantization of an electromagnetic field is presented in the form

$$A_\mu(x)=\int \frac{d^3k}{(2\pi)^3}\frac{1}{\sqrt{2k^0}}\{a_\mu(\mathbf{k})e^{ikx}+a_\mu^+(\mathbf{k})e^{-ikx}\}. \tag{A1}$$

Here $\mu,\nu,\ldots=0,1,2,3$; $kx\equiv k_\mu x^\mu=-k^0x^0+\mathbf{k}\cdot\mathbf{x}$, $k^0=|\mathbf{k}|$ and $\{a_\mu(\mathbf{k}),\ a_\mu^+(\mathbf{k})\}$ are generators of an associative involutory algebra $\mathscr{A}$ with an identity element (see Sec. 2). The nonzero commutation relations between these generators have the form

$$[a_\mu(\mathbf{k}),a_\mu^+(\mathbf{p})]=(2\pi)^3\eta_{\mu\nu}\delta^{(3)}(\mathbf{k}-\mathbf{p}). \tag{A2}$$

One can see from the expansion (A1) that the set of elements $\partial_\mu A^\mu(x)$ is linearly equivalent to the set of elements $k^\mu a_\mu(\mathbf{k})$ and $k^\mu a_\mu^+(\mathbf{k})$ from the algebra $\mathscr{A}$. Let $a_i^T(\mathbf{k})$ be two independent elements (for fixed $\mathbf{k}$) satisfying the conditions

$$\sum_{i=1}^{3}k_i a_i^T(\mathbf{k})=0,$$

$$[a_i^T(\mathbf{k}),\ a_j^{T+}(\mathbf{p})]=(2\pi)^3\left(\delta_{ij}-\frac{k_ik_j}{\mathbf{k}^2}\right)\delta^{(3)}(\mathbf{k}-\mathbf{p}). \tag{A3}$$

From Eqs. (A1) and (A2) follow the commutation relations ($F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$)

$$[F_{\mu\nu}(x),\ k^\lambda a_\lambda(\mathbf{k})]=[F_{\mu\nu}(x),\ k^\lambda a_\lambda^+(\mathbf{k})]=0. \tag{A4}$$

$$[k^\mu a_\mu(\mathbf{k}),\ p^\nu a_\nu^+(\mathbf{p})]=0. \tag{A5}$$

It is obvious that

$$[a_i^T,k^\mu a_\mu(\mathbf{k})]=[a_i^T,\ k^\mu a_\mu^+(\mathbf{k})]=0. \tag{A6}$$

Dirac quantization presupposes that the condition

$$a_i^T(\mathbf{k})|0\rangle=0 \tag{A7}$$

is imposed on the ground state and the conditions

$$k^\mu a_\mu(\mathbf{k})|\rangle=0,\quad k^\mu a_\mu^+(\mathbf{k})|\rangle=0. \tag{A8}$$

are imposed on all states. As a result of Eqs. (A5) and (A6) the conditions (A7) and (A8) are algebraically compatible. The states satisfying the conditions (A8) are called physical. The Fock space of all physical states is constructed with the aid of the creation operators $a_i^{T+}(\mathbf{k})$ from the ground state satisfying the conditions (A7) and (A8). As a result of Eq. (A6) any state of the Fock space constructed satisfies the conditions (A8). Following the terminology introduced in Sec. 2, this Fock space is designated by the symbol V, the set of elements $\{a_i^T,\ a_i^{T+},\ k^\mu a_\mu(\mathbf{k}),\ k^\mu a_\mu^+(\mathbf{k})\}$ is a system of generators of the subalgebra $\mathscr{B}$ and the set of elements $\{k^\mu a_\mu(\mathbf{k}),\ k^\mu a_\mu^+(\mathbf{k})\}$ is a system of generators of the subalgebra $\mathscr{N}$.

Let $k_-^\mu=(-k^0,\ \mathbf{k})$. We find from Eq. (A2)

$$[k_-^\mu a_\mu(\mathbf{k}),\ p^\nu a_\nu^+(\mathbf{p})]=2\mathbf{k}^2(2\pi)^2\delta^{(3)}(\mathbf{k}-\mathbf{p}). \tag{A9}$$

The relations (A4) and (A9) signify that the observables $F_{\mu\nu}$ do not depend on the generators $\{k_-^\mu a_\mu(\mathbf{k}),\ k_-^\mu a_\mu^+(\mathbf{k})\}$, but rather they are linear combinations of the generators of the subalgebra $\mathscr{B}$. Therefore all matrix elements of the form $\langle\Lambda|F_{\mu\nu}|\Sigma\rangle$, where $|\Lambda\rangle,\ |\Sigma\rangle\in V$, are determined.

We note that as a result of Eqs. (A3) and (A7) the scalar product in the space V is positive-definite provided that $\langle 0|0\rangle=1$. We call attention to the fact that the action of the generators $k_-^\mu a_\mu(\mathbf{k})$ and $k_-^\mu a_\mu^+(\mathbf{k})$ on the physical states is not determined in Dirac quantization, and therefore these generators of the algebra $\mathscr{A}$ are not linear operators in the space of physical states V.

In closing, we call attention to an analogy between the generators $\{k_-^\mu a_\mu(\mathbf{k}),\ k_-^\mu a_\mu^+(\mathbf{k})\}$ and $\{k^\mu a_\mu(\mathbf{k}),\ k^\mu a_\mu^+(\mathbf{k})\}$ in quantum electrodynamics and the generators $\{\alpha_n^{(+)}\}$ and $\{\alpha_n^{(-)}\}$ in string theory, respectively.

*)E-mail: vergeles@itp.ac.ru

[1] R. Jackiw, E-print archive, gr-qc/9612052.
[2] D. Gangemi and R. Jackiw, Phys. Lett. B **337**, 271 (1994); Phys. Rev. D **50**, 3913 (1994); D. Amati, S. Elitzur, and E. Rabinovici, Nucl. Phys. B **418**, 45 (1994); D. Louis-Martinez, J. Gegenberg, and G. Kunstatter, Phys. Lett. B **321**, 193 (1994); E. Benedict, Phys. Lett. B **340**, 43 (1994); T. Strobl, Phys. Rev. D **50**, 7346 (1994).
[3] C. Callan, S. Giddings, J. Harvey, and A. Strominger, Phys. Rev. D **45**, 1005 (1992).
[4] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory*, Cambridge University Press, Cambridge, 1987 [Russian translation, Mir, Moscow, 1990].
[5] P. A. M. Dirac, *Lectures on Quantum Field Theory*, Yeshiva University Press, N.Y., 1967 [Russian translation, Mir, Moscow, 1971].
[6] S. Lang, *Algebra*, Addison-Wesley, Reading, MA, 1965 [Russian translation, Mir, Moscow, 1968].
[7] N. Bourbaki, *Spectral Theory* [Russian translation, Mir, Moscow, 1972].

Translated by M. E. Alferieff

Ionization of atoms in electric and magnetic fields and the imaginary time method

V. S. Popov

Institute of Theoretical and Experimental Physics, 117259 Moscow, Russia

B. M. Karnakov*) and V. D. Mur

Moscow Engineering Physics Institute (Technical University), 115409 Moscow, Russia

(Submitted 25 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1579–1605 (May 1998)

A semiclassical theory is developed for the ionization of atoms and negative ions in constant, uniform electric and magnetic fields, including the Coulomb interaction between the electron and the atomic core during tunneling. The case of crossed fields (Lorentz ionization) is examined specially, as well as the limit of a strong magnetic field. Analytic equations are derived for arbitrary fields $\mathscr{E}$ and $\mathscr{H}$ that are weak compared to the characteristic intraatomic fields. The major results of this paper are obtained using the ''imaginary time'' method (ITM), in which tunneling is described using the classical equations of motion but with purely imaginary ''time.'' The possibility of generalizing the ITM to the relativistic case, as well as to states with nonzero angular momentum, is pointed out. © *1998 American Institute of Physics.* [S1063-7761(98)00405-3]

INTRODUCTION AND STATEMENT OF THE PROBLEM

The problem of the hydrogen atom in electric and magnetic fields is of fundamental significance for atomic physics and is often encountered in applications. An extensive literature has been devoted to this topic[1–30] (see the references in Refs. 8, 9, 19, and 20 as well). The properties of the energy spectrum of atomic hydrogen and other atoms in strong external fields have been discussed by many authors,[9–16] but the level widths $\Gamma_n=\hbar w_n$ have usually been neglected.

Problems associated with the ionization of atoms and ions in strong fields became especially timely after the development of lasers. In the 1960's a semiclassical theory of ionization in an electric field $\mathscr{E}$ was developed for neutral atoms[1–6] and negative ions such as H^-, J^-, etc.[7,8] (the first of these problems is more complicated, owing to the need to allow for the Coulomb interaction between the ejected electron and the atomic core). The semiclassical theory yields analytic equations for the ionization probability w which are asymptotically exact in the limit of ''weak'' fields (see the conditions (1.2) below). Currently, numerical calculations[11,15–28] are also available in which the positions and widths of the atomic levels have been calculated in pure electric[17–24] and magnetic[25–27] fields, as well as in parallel[11,15,28] and mutually perpendicular[26,27] $\mathscr{E}$ and $\mathscr{H}$ fields. Despite the obvious value of such calculations (as applied to selected magnitudes of the fields and to specific atoms), they cannot replace an analytic theory.

The effect of the magnetic field $\mathscr{H}$ on the ionization probability $w(\mathscr{E},\mathscr{H})$ has been studied,[29–31] but it has been possible to consider only the case of negative ions, where there is no Coulomb interaction at large distances from an atom. Both the exponential factor[29] in the probability w and the multiplicative factor[30,31] have been calculated (albeit the latter only in a few special cases).

Including the Coulomb interaction in problems of this kind presents great difficulty and, for example, it has not yet been fully[1)] accounted for in the theory of multiphoton ionization of atoms.[32,3,4] For the ionization of an atomic level by constant and uniform $\mathscr{E}$ and $\mathscr{H}$ fields, this type of calculation can be carried out analytically if the tunneling transition of the electron is described using the ''imaginary time'' method (ITM).[3,36] The results of this calculation are discussed below.

This paper is organized as follows. In Sec. 1 the basic equations for the ionization probability $w(\mathscr{E},\mathscr{H})$ are introduced and the limiting cases of strong and weak magnetic fields are discussed. Section 2 is devoted to the special case $\mathscr{E}\perp\mathscr{H}$, which includes the particular case of Lorentz ionization,[30,37] which arises during the motion of atoms in a constant magnetic field. In Sec. 3, a simple asymptotic theory is developed for the case of a strong magnetic field $\gamma\gg 1$ (see Eq. (1.7) below). In Sec. 4, equations are introduced for the asymptotic coefficient C_κ, in terms of which the ionization probability of the atomic s-level can be expressed. In Sec. 5, the main computational technique, the ITM, is described. It is shown that with this method it is easy to obtain the basic equations in this paper. A brief review of the results obtained in this paper is contained in Sec. 6.

In Appendix A we present an explicit expression for the Green function of an electron in $\mathscr{E}$ and $\mathscr{H}$ fields with arbitrary magnitudes and directions, and derive an integral equation for the energy of the quasistationary state in this case $(Z=0)$, when there is no Coulomb interaction at large distances. Some auxiliary equations and asymptotic dependences are given in Appendix B and the details of the calculation for the case $\gamma\gg 1$ are discussed in Appendix C. Some of the results in this paper have been announced previously.[38,39]

We dedicate this paper to the memory of Mikhaĭl Vasil'evich Terent'ev (1935–1996), an outstanding theoreti-

TABLE I. Parameters of atoms and ions.

Atom or ion	κ	C_κ	Atom or ion	κ	C_κ
^{1}H($1s$)	1.000	1.00	^{24}Cr($4s$)	0.705	1.10
H$^-$($1s$)	0.236	1.15	^{25}Mn($4s$)	0.739	1.15
^{2}He($1s$)	1.344	0.99	^{26}Fe($4s$)	0.762	1.15
He$^+$($1s$)	2.000	1.00	Fe$^+$($4s$)	1.091	1.10
He$^-$($2p$)	0.075	–	Fe$^-$($4s$)	0.171	1.12
^{3}Li($2s$)	0.630	1.07	^{29}Cu($4s$)	0.754	1.08
Li$^+$($1s$)	2.358	0.99	Cu$^-$($4s$)	0.301	1.09
Li$^-$($2s$)	0.212	1.09	^{37}Rb($5s$)	0.554	0.93
^{4}Be($2s$)	0.828	1.12	Rb$^-$($5s$)	0.189	0.92
Be$^+$($2s$)	1.157	0.97	^{38}Sr($5s$)	0.645	1.05
^{5}B($2p$)	0.781	–	^{47}Ag($5s$)	0.746	1.01
B$^+$($2s$)	1.360	1.8	Ag$^-$($5s$)	0.309	1.2
^{11}Na($3s$)	0.615	1.04	^{55}Cs($6s$)	0.535	0.93
Na$^-$($3s$)	0.201	1.1	^{56}Ba($6s$)	0.619	1.07
^{12}Mg($3s$)	0.750	1.12	Ba$^+$($6s$)	0.857	1.00
Mg$^+$($3s$)	1.051	1.03	^{79}Au($6s$)	0.823	1.10
^{13}Al$^+$($3s$)	1.176	1.1	^{80}Hg($6s$)	0.876	1.22
^{19}K($4s$)	0.565	0.95	Hg$^+$($6s$)	1.174	1.30
K$^-$($4s$)	0.192	1.03	^{90}Th($7s$)	0.67	1.11
^{20}Ca($4s$)	0.670	1.05	^{92}U($7s$)	0.674	1.08
Ca$^+$($4s$)	0.934	0.97			

Note: Parameters for outer (valence) electrons in the states indicated in parentheses are listed in the table.

cal physicist and one of the originators of the imaginary time method, with whom we had valuable conversations and who taught us much.

1. EFFECT OF A MAGNETIC FIELD ON THE IONIZATION PROBABILITY FOR ATOMS AND NEGATIVE IONS

Let $E_0=-\kappa^2 me^4/2\hbar^2$ be the energy of the unperturbed atomic level with orbital angular momentum $l=0$ (we neglect spin in the following), and ϵ and h be the reduced values of the external fields:

$$\epsilon=\mathscr{E}/\kappa^3\mathscr{E}_a, \quad h=\mathscr{H}/\kappa^2\mathscr{H}_a. \tag{1.1}$$

where $\mathscr{E}_a=m^2e^5/\hbar^4=5.142\cdot 10^9$ V/cm and $\mathscr{H}_a=m^2ce^3/\hbar^3=2.350\cdot 10^9$ G are the atomic units for the field strengths (in the following, as a rule $\hbar=e=m=1$, but in the final equations we restore the dimensionality of the quantities contained in them). The ratio $\mathscr{H}/\mathscr{E}$, as well as the angle θ between the fields, can be arbitrary. The condition for applicability of the semiclassical approximation in this case is

$$\epsilon\ll 1, \quad h\ll 1. \tag{1.2}$$

For the ground states of the neutral atoms $\kappa\approx 1$ (see Table I), but for highly excited (Rydberg) states and negative ions it can be much smaller than unity. In this case, values $\epsilon,h\sim 1$ are attained even for fields substantially lower than the atomic fields (this applies to semiconductors as well[40]).

For calculating the ionization probabilities w we shall use the ITM. In this method electron tunneling is described using subbarrier trajectories that satisfy the classical equations of motion, but with imaginary ''time.'' The imaginary part of the reduced action W (see Eq. (5.1) below) calculated along this trajectory determines the tunneling probability, i.e., (in this case) the ionization probability of a bound state:

$$w(\mathscr{E},\mathscr{H})\propto\exp\left\{-\frac{2}{\hbar}\,\mathrm{Im}\;W(0,t_0)\right\}, \tag{1.3}$$

where t_0 is the initial time for subbarrier motion, while $t=0$ is the time of particle escape from under the barrier. We leave a detailed discussion of the ITM to Sec. 5. There it will also be shown that in the case of constant $\mathscr{E}$ and $\mathscr{H}$ fields, the Coulomb interaction of an electron with the atomic core (charge Z) can be taken into account using a particular variant of perturbation theory, so that it is enough to determine the subbarrier trajectory of an electron for $Z=0$, i.e., in effect for a δ-potential. The extremal trajectory which minimizes the imaginary part of the action determines the most probable tunneling path for a particle from a bound state to the continuum, and is found using the classical equations of motion with boundary conditions

$$\mathbf{r}(t_0)=0, \quad \dot{\mathbf{r}}^2(t_0)=-\kappa^2, \quad \mathrm{Im}\;\mathbf{r}(0)=\mathrm{Im}\;\dot{\mathbf{r}}(0)=0. \tag{1.4}$$

Here we note their intuitive significance, referring the reader to Ref. 36 for details. The first two conditions mean that at the initial time the electron is already beyond the confines of the atomic core (i.e., $|V(r_0)|\ll\kappa^2$), but the distortion of the wave function caused by the external $\mathscr{E}$ and $\mathscr{H}$ fields can still be neglected.[2)] The latter condition means that the most probable (extremal) trajectory at $t=0$ becomes real and subsequently describes the motion of the particle at infinity, even in the classically allowed region. Setting $\tau=-i\omega_c t$, we find the desired trajectory:[3)]

$$x = i\,\frac{|e|\mathscr{E}}{m\omega_c^2}\left(\tau - \tau_0\,\frac{\sinh\,\tau}{\sinh\,\tau_0}\right)\sin\,\theta,$$

$$y = \frac{e\mathscr{E}}{m\omega_c^2}\,\frac{\tau_0}{\sinh\,\tau_0}\,(\cosh\,\tau_0 - \cosh\,\tau)\sin\,\theta,$$

$$z = \frac{e\mathscr{E}}{2m\omega_c^2}\,(\tau_0^2 - \tau^2)\cos\,\theta \tag{1.5}$$

$(-\tau_0 \leqslant \tau \leqslant \tau_0)$. The second of the conditions (1.4) yields an equation for the initial time of subbarrier motion $t_0 = -i\omega_c^{-1}\tau_0$:

$$\tau_0^2 - \sin^2\,\theta(\tau_0\,\coth\,\tau_0 - 1)^2 = \gamma^2, \tag{1.6}$$

or

$$\frac{\tanh\,\tau_0}{\tau_0} = \frac{\sin\,\theta}{\sin\,\theta + \sqrt{\tau_0^2 - \gamma^2}}. \tag{1.6′}$$

Here

$$\gamma = \frac{|\omega_c|}{\omega_t} = \frac{\kappa\mathscr{H}}{c\mathscr{E}} = \frac{h}{\epsilon}, \tag{1.7}$$

$\omega_c = e\mathscr{H}/mc$ is the Larmor or cyclotron frequency, and $\omega_t = \mathscr{E}/\kappa$ is the characteristic tunneling frequency in the electric field. In this case, the imaginary tunneling time of the particle is

$$T_t = \int_0^b \frac{dx}{|p(x)|} = \frac{\kappa}{\mathscr{E}} \equiv \frac{1}{\omega_t},$$

where $|p(x)| = \sqrt{\kappa^2 - 2\mathscr{E}x}$ and $b = \kappa^2/2\mathscr{E}$ is the barrier width. Note that the parameter γ is analogous to the Keldysh parameter[32] which shows up in the theory of the multiphoton ionization of atoms by a variable electric field. It appears because in this problem there are two frequencies, ω_t and ω_c, which can have an arbitrary ratio. Then, we have

$$\gamma = \sqrt{\frac{I}{I_0}}\,\frac{\mathscr{H}}{\mathscr{H}_a}\left(\frac{\mathscr{E}}{\mathscr{E}_a}\right)^{-1}, \tag{1.7′}$$

where $I = \kappa^2/2$ is the binding energy of the level and $I_0 = 13.6$ eV is the ionization potential of the hydrogen atom. (the parameter κ includes the shift of the level in an external field).

It is clear from Eq. (1.6′) that there is a unique (real) solution $\tau_0 = \tau_0(\gamma,\theta)$, with $\tau_0 \geqslant \gamma$. If on the other hand $\gamma > 1$, then the following stronger constraint holds:

$$\tau_0 > \tilde{\tau}_0 \equiv \frac{\sin^2\,\theta + \gamma^2}{\sin^2\,\theta + \sqrt{\sin^2\,\theta + \gamma^2\,\cos^2\,\theta}}, \tag{1.8}$$

which becomes asymptotically exact for $\gamma \gg 1$ (see Fig. 1).

In Eqs. (1.5) and (1.6) and in the subsequent equations, it is assumed that the magnetic field is directed along the z axis, the x axis is perpendicular to the $(\mathscr{E},\mathscr{H})$ plane, and θ is the angle between $\mathscr{E}$ and $\mathscr{H}$. Substituting Eq. (1.5) in Eq. (1.3) yields the exponential factor in the ionization probability. In order to determine the preexponential factor, we also have to include the contribution from a beam of near-extremal subbarrier trajectories. As a result of the calculations, we arrive at the following formula:

$$w(\mathscr{E},\mathscr{H}) = \frac{me^4\kappa^2}{\hbar^3}\cdot 2^{2\eta}|C_\kappa|^2\left(\frac{\hbar^4\mathscr{E}}{m^2e^5\kappa^3}\right)^{1-2\eta} \times PQ^\eta\,\exp\left\{-\frac{2m^2e^5\kappa^3}{3\hbar^4\mathscr{E}}\,g(\gamma,\theta)\right\}. \tag{1.9}$$

Here C_κ is the asymptotic coefficient for the normalized wave function of the initial state $(l=0)$:

$$\psi_0(r) \approx C_\kappa\sqrt{\kappa^3/\pi}\;e^{-\kappa r}(\kappa r)^{\eta-1}, \quad r \gg \kappa^{-1} \tag{1.10}$$

(in particular, $C_\kappa = 1$ for the $1s$- and $2s$-states of the hydrogen atom and $C_\kappa = 1/\sqrt{2}$ for the level in a three dimensional delta function potential),

$$g(\gamma,\theta) = \frac{3}{2}\,\beta\left(1 - \frac{\sqrt{\beta^2-1}}{\gamma}\,\sin\,\theta - \frac{1}{3}\,\beta^2\,\cos^2\,\theta\right), \tag{1.11}$$

$\beta = \tau_0/\gamma \geqslant 1$ and the Coulomb factor has the form ($\eta = Z/\kappa$ is the Sommerfeld parameter)

$$Q(\gamma,\theta) = \exp\left\{2\left(\ln\frac{\beta}{2} + \int_0^{\tau_0} d\tau\left[\frac{\gamma}{\xi(\tau)} - \frac{1}{\tau_0 - \tau}\right]\right)\right\}, \tag{1.12}$$

where

$$\xi(\tau) = \left\{\frac{1}{4}\,(\tau_0^2 - \tau^2)^2\,\cos^2\,\theta + \tau_0^2\left[\left(\frac{\cosh\,\tau_0 - \cosh\,\tau}{\sinh\,\tau_0}\right)^2 - \left(\frac{\sinh\,\tau}{\sinh\,\tau_0} - \frac{\tau}{\tau_0}\right)^2\right]\sin^2\,\theta\right\}^{1/2} \tag{1.12′}$$

and, finally, the preexponential factor is

$$P(\gamma,\theta) = \frac{\gamma^2}{\tau_0}\left[\left(\frac{\sinh\,2\tau_0}{2\tau_0} + \tau_0\,\coth\,\tau_0 - 2\right) \times \sin^2\,\theta + \sinh^2\,\tau_0\,\cos^2\,\theta\right]^{-1/2}. \tag{1.13}$$

The simplest way of obtaining these formulas is to use the ITM; this is discussed in Sec. 5. The preexponential factor $P(\gamma,\theta)$ can also be calculated independently by solving an integral equation for the (complex) energy of the quasistationary state, $E = E_r - i\Gamma/2$:

$$\kappa = \kappa_0 + \kappa\sqrt{\frac{h}{4\pi i}}\int_0^\infty \frac{du}{\sqrt{u^3}} \times\left\{\frac{u}{\sin\,u}\,\exp[-i\Lambda\varphi(u)] - \exp(i\Lambda\gamma^2 u)\right\} \tag{1.14}$$

(see Appendix A), where $\kappa_0 = \sqrt{-2E_0}$ and $\kappa = \sqrt{-2E}$ are the parameters for a free $(\mathscr{E} = \mathscr{H} = 0)$ atom and for an atom in external fields,

$$\Lambda=\epsilon^2 h^{-3}, \quad \varphi(u)=[\gamma^2+(1-u\cot u)\times\sin^2\theta]u+\frac{1}{3}u^3\cos^2\theta. \tag{1.15}$$

Equation (1.14) defines both the level shift E_r-E_0 and its width Γ, and is obtained if we impose on the function (A13) the boundary condition at $r\to 0$ corresponding to a δ-potential, i.e., to the approximation of zero-range forces (in Appendix A we obtain Eq. (A18), which includes a correction for the effective radius). If $\Lambda\gg 1$, then the integral in Eq. (1.14) can be calculated by the method of steepest descents, and this leads to Eq. (1.9) with $\eta=0$ in the preexponential factor (1.13). The calculation of the Coulomb factor (1.12) is discussed below in Sec. 5.

We now discuss the results.

a) In the absence of a magnetic field ($\gamma=0$), the equations can be simplified: $g(\gamma,\theta)=P=Q\equiv 1$ and

$$w(\mathscr{E})=2^{2\eta}\kappa^2|C_\kappa|^2\epsilon^{1-2\eta}\exp(-2/3\epsilon), \tag{1.16}$$

in agreement with the results of Refs. 1–3 and 7 for an s level.

b) In the case of parallel fields, we have $\tau_0=\gamma$,

$$g(\gamma,0)=Q(\gamma,0)\equiv 1, \quad P(\gamma,0)=\gamma/\sinh\gamma. \tag{1.17}$$

c) For $\gamma\ll 1$ (i.e., at low magnetic fields) and arbitrary angles θ, we can use an expansion of these functions, which is given in Appendix B (see Eqs. (B1)–(B4)).

d) In the opposite case of $\gamma\to\infty$ the functions $g(\gamma,\theta)$ and $Q(\gamma,\theta)$ approach constant limits (if $\theta<\pi/2$) and the preexponential factor $P(\gamma,\theta)$ falls off exponentially; see the asymptotes (B6)–(B9) in Appendix B. The case $\theta=\pi/2$ is special and will be discussed in the following section.

e) If $\theta\neq 0$, then $g(\gamma,\theta)>1$, so the magnetic field reduces the ionization probability, stabilizing the level. This is related to the fact that because of the Lorentz force, the subbarrier trajectory of the electron is ''twisted'' and the width of the barrier increases (an intuitive explanation of this is provided by the ITM; see Sec. 5). On the other hand, including the Coulomb interaction greatly increases the probability of ionizing a neutral atom compared to the case of a negative ion (for equal binding energies $\kappa^2/2$ and comparable values of the constants C_κ^2). This is explained by the fact that in a Coulomb field the electron density at the ''edge'' of the atom is higher; see Eqs. (5.8) and (5.9) below.

f) The Coulomb and preexponential factors in Eq. (1.9) are corrections (both of the same order of magnitude $\sim\ln\epsilon$ in the small parameter $\epsilon\to 0$) to the leading term, which is proportional to ϵ^{-1}, in the exponent. For this reason, $Q(\gamma,\theta)$ and $P(\gamma,\theta)$ can be calculated independently of one another.

g) The leading (exponential) factor in Eq. (1.9) has been calculated before (see Eqs. (35) and (36) in Ref. 29); this yielded a more cumbersome expression than Eq. (1.11) for $g(\gamma,\theta)$, but reduces to it when Eq. (1.6) is used. Note that the coefficient of γ^2 in the expansion of the preexponential factor P (which is independent of the angle θ according to Eq. (B2)) is the same as that calculated previously for the special cases $\theta=0$[31] and $\pi/2$.[30]

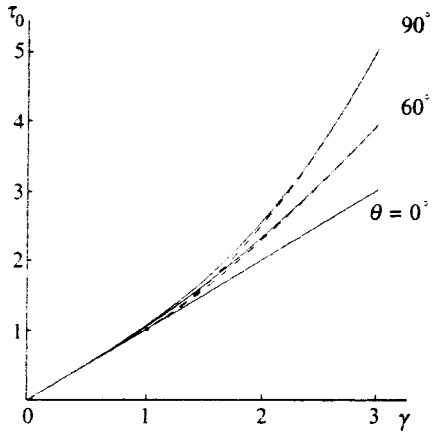

FIG. 1. τ_0 as a function of the parameter γ according to Eq. (1.6). The angles θ between the electric and magnetic fields are indicated next to the curves. The dashed curves correspond to the approximation $\tilde{\tau}_0$ from Eq. (1.8); here $\tilde{\tau}_0=1$ for $\gamma=1$ and arbitrary θ.

The results of some numerical calculations using these formulas are shown in Figs. 1 and 2. Note that $\tau_0=\tau_0(\gamma,\theta)$ has a simple physical significance: $t_0=-i\omega_c^{-1}\tau_0$, where t_0 is the time (purely imaginary) of subbarrier motion. The functions g, P, and Q are plotted as functions of γ in Figs. 2a–c.

With rising magnetic field, the barrier width increases, while the ionization probability decreases. We introduce a ''stabilization factor'' S, which accounts for the suppression of the decay of a bound state by a magnetic field:

$$w(\mathscr{E},\mathscr{H})=Sw(\mathscr{E},0). \tag{1.18}$$

For parallel fields this factor can be found analytically:

$$S(\gamma,0)=\frac{\gamma}{\sinh\gamma}=\begin{cases}1-\frac{1}{6}\gamma^2+\frac{7}{360}\gamma^4+\ldots & \gamma\to 0,\\ 2\gamma e^{-\gamma}+O(e^{-3\gamma}), & \gamma\to\infty,\end{cases} \tag{1.19}$$

and, in this case, it is the same for neutral atoms and negative ions. In the general case, we have

$$S=PQ^\eta\exp\left\{-\frac{2}{3\epsilon}[g(\gamma,\theta)-1]\right\}, \tag{1.20}$$

so that for $\gamma\to 0$ we obtain

$$S=1-s_1\gamma^2+\ldots, \quad s_1=\frac{1}{6}+\frac{2}{9}(0.1\epsilon^{-1}-\eta)\sin^2\theta. \tag{1.21}$$

As a rule, the coefficient s_1 is positive (in particular, for all negative ions). The stabilization factor falls off rapidly as the magnetic field increases, especially when $\gamma>1$ (Fig. 3). The dependence of S on the Coulomb parameter η becomes significant for $\gamma>2$. The preexponential factor P sharply reduces the probability w if $\gamma\gtrsim 10$ (Fig. 2b).

2. THE CASE $\mathscr{E}\perp\mathscr{H}$

The case of mutually perpendicular fields is special, both formally (the asymptotes of the functions in Eq. (1.9) have a

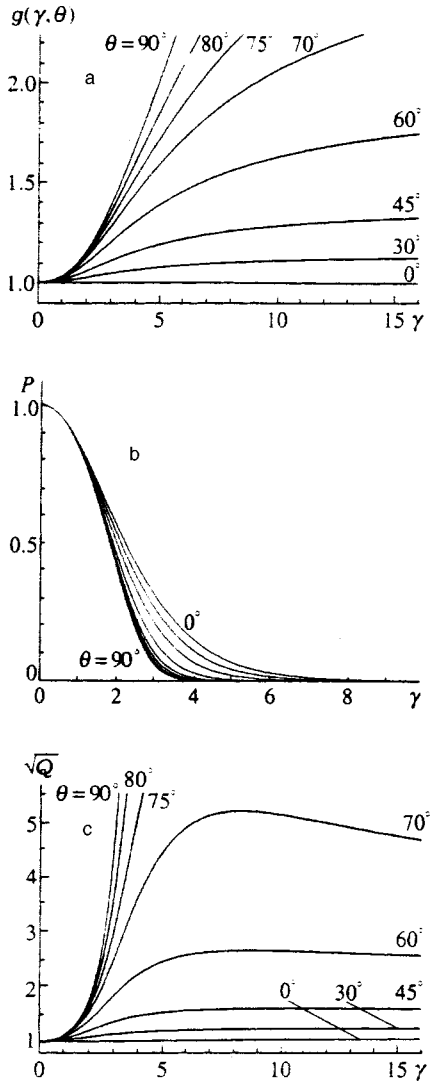

FIG. 2. The functions g (a), P (b), and Q (c) in Eq. (1.9). In b the curves correspond to the same angles θ as in a. Because of the strong dependence of $Q(\gamma,\theta)$ on γ, in c we have introduced $\sqrt{Q}$.

different form for large γ) and physically: a field configuration of this type shows up in the rest frame of an atom moving in a constant magnetic field (Lorentz ionization[37]). The equations take the form

$$g(\gamma)=\frac{3}{2}\,\beta^3\left[1-\left(\coth\,\tau_0-\frac{1}{\tau_0}\right)\coth\,\tau_0\right],$$

$$P(\gamma)=\frac{\gamma}{\beta}\left[\left(\frac{\sinh\,\tau_0}{\tau_0}+\frac{\tau_0}{\sinh\,\tau_0}\right)\cosh\,\tau_0-2\right]^{-1/2}, \quad (2.1)$$

and

$$Q(\gamma)=\frac{1}{4}\,\beta^2\,\exp\left(2\int_0^{\tau_0}d\tau\left\{\beta^{-1}\left[\left(\frac{\cosh\,\tau_0-\cosh\,\tau}{\sinh\,\tau_0}\right)^2-\left(\frac{\sinh\,\tau}{\sinh\,\tau_0}-\frac{\tau}{\tau_0}\right)^2\right]^{-1/2}-\frac{1}{\tau_0-\tau}\right\}\right), \quad (2.2)$$

where $\tau_0(\gamma)$ is determined from the equation $\tau_0=\tau_0/(1+\sqrt{\tau_0^2-\gamma^2})$, and $\beta=[1-\coth\,\tau_0-1/\tau_0)^2]^{-1/2}$.

For $\gamma\ll 1$ (weak magnetic field),

$$g(\gamma)=1+\frac{1}{30}\,\gamma^2+\frac{11}{7560}\,\gamma^4+\frac{53}{2\,041\,200}\,\gamma^6-\frac{2507}{538\,876\,800}\,\gamma^8+\ldots, \quad (2.3)$$

and

$$P(\gamma)=1-\frac{1}{6}\,\gamma^2-\frac{1}{3240}\,\gamma^4+\ldots, \quad Q(\gamma)=1+\frac{2}{9}\,\gamma^2+\ldots\,. \quad (2.4)$$

On the other hand, for $\gamma\gg 1$, we have (see Appendices B and C)

$$g(\gamma)=\frac{3}{8}\,\gamma(1+\gamma^{-2})^2[1-c_1\epsilon^{-\gamma^2}+O(e^{-2\gamma^2})], \quad (2.5)$$

and

$$P(\gamma)=c_2e^{-\gamma^2/2}\gamma\left(1-\frac{1}{2\gamma^2}+\ldots\right),$$

$$Q(\gamma)=c_3e^{\pi\gamma}\gamma^{-2}[1+O(\gamma^{-1})], \quad (2.6)$$

where the c_i are numerical coefficients: $c_1=2e^{-1}=0.7358$, $c_2=2^{3/2}e^{-1/2}=1.716$, and $c_3=0.0106$. Thus, $g(\gamma)\propto\gamma$ for $\gamma\to\infty$, while for angles $\theta<\pi/2$ this function approaches a constant; asymptotically, of the functions P and Q also differ substantially.

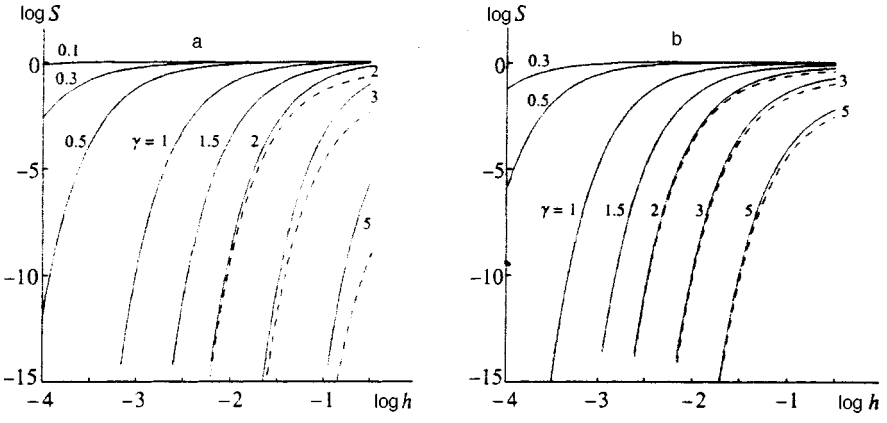

FIG. 3. The stabilization factor S for $\theta=90°$ (a) and 45° (b): the smooth curves are for the ground state of the hydrogen atom ($\kappa=\eta=1$) and the dashed curves, for a negative ion with $\kappa=1$ and $\eta=0$. The values of γ are indicated on the curves ($h=\mathscr{H}/\mathscr{H}_a$).

TABLE II. Approximations to $g(\gamma)$.

γ	A	B	δ	C	δ
1.0	1.0348090	1.03479	[−2.0(−3)]	1.09399	[5.7]
		1.0348097	[6.3(−5)]		
1.5	1.0825046	1.08237	[−0.013]	1.08260	[8.7(−3)]
		1.08254	[3.5(−3)]		
2.0	1.1564487	1.15661	[0.014]	1.15608	[−0.032]
		1.15708	[0.055]		
2.5	1.2597242	1.26517	[0.43]	1.25971	[−1.3(−3)]
		1.26441	[0.37]		

Note: The table contains values of $g(\gamma,\theta)$ for $\theta=\pi/2$: A) numerical calculation; B) according to Eq. (2.3) with terms up to γ^4 and γ^8, inclusive, retained in the upper and lower lines, respectively; C) the asymptote (2.5). The relative errors δ in the corresponding approximations (percent) are given in square brackets; the order of magnitude is indicated in parentheses, i.e., $a(b)\equiv a\cdot 10^b$.

A remarkable feature of these expansions is that their domains of applicability ''stretch down'' to $\gamma\sim 1$, where these asymptotes merge smoothly into one another. This is illustrated in Table II, which lists values of $g(\gamma)$ calculated using the approximate equations (2.3) and (2.5) in columns B and C, respectively, as well as the results of a numerical calculation (column A). It is clear that for $\gamma\lesssim 2$, $g(\gamma)$ can be calculated using Eq. (2.3), and that for $\gamma>1.5$ it can be calculated using the asymptotic expansion (2.5) without having to solve a transcendental equation. Although the functions $P(\gamma)$ and $Q(\gamma)$ vary more rapidly than $g(\gamma)$ when $\gamma\gg 1$, the probability w is most sensitive to the changes in $g(\gamma)$ itself, since this function appears in the exponent of Eq. (1.9), and with a large coefficient $2/3\epsilon$ at that. Here the Coulomb and preexponential factors are parametrically small relative to the leading asymptotic term $\exp\{-(2/3\epsilon)g(\gamma)\}$; to exponential accuracy, we have

$$w(\gamma)\approx\exp\left\{-\frac{h}{4\epsilon^2}\left[1+2(h-2\pi\eta\epsilon)+\ldots\right]\right\},$$

$$\gamma=\frac{h}{\epsilon}\gg 1, \tag{2.7}$$

where the term $2h$ in square brackets comes from the preexponential factor $P(\gamma)$, while the term proportional to η comes from the Coulomb factor $Q(\gamma)$.

It is interesting to note that for the ground state of the hydrogen atom ($\kappa=\eta=1$), the overall factor PQ in front of the exponential in Eq. (1.9) is close to unity for $0\leqslant\gamma\lesssim 4$ (Fig. 4), although $P(\gamma)$ and $Q(\gamma)$ individually vary within this range of γ by more than two orders of magnitude (e.g., $P(4)=2.23\times 10^{-3}$, while $Q(4)=227.3$). Here essentially all the dependence of the ionization probability on γ is determined by the exponential factor (if $\eta\neq 1$, then this is no longer true).

Note that in the case $\theta=\pi/2$ and $\gamma>1$, although the ionization probability w is small (especially in view of the linear growth in $g(\gamma)$ as $\gamma\to\infty$), it does not go identically to zero, notwithstanding the claim made in Ref. 30. The equations in Ref. 30 for w in the region $\gamma<1$ are also wrong, except for the expansion (9) for $\gamma\to 0$.

3. ASYMPTOTIC BEHAVIOR FOR $\gamma\gg 1$

In this case, which is analogous to the antiadiabatic ($\omega\gg\omega_t$) approximation in the theory of multiphoton ionization, we can develop a simpler asymptotic theory (this approach may also be useful for a number of other problems.)

The equations for the subbarrier trajectory of the electron take the form

$$x=ia\left(\frac{\tau}{\tau_0}-e^{\tau-\tau_0}\right)\sin\theta,\quad y=-a(1-e^{\tau-\tau_0})\sin\theta,$$

$$z=-\frac{a}{2\tau_0}(\tau_0^2-\tau^2)\cos\theta, \tag{3.1}$$

where $a=\mathscr{E}\tau_0\omega_c^{-2}$, $\tau=-i\omega_c t$, and $\tau_0\gg 1$. Here

$$\dot{\mathbf{r}}^2=-\frac{\kappa^2}{\gamma^2}\{\tau^2\cos^2\theta-(1-2\tau_0 e^{\tau-\tau_0})\sin^2\theta\}. \tag{3.2}$$

The initial condition $\dot{\mathbf{r}}^2(\tau_0)=-\kappa^2$ yields the approximation (1.8) for τ_0, which is highly accurate for $\gamma\gtrsim 1$. Thus, in the case of parallel fields $\tilde{\tau}_0=\tau_0=\gamma$, while for $\theta=\pi/2$ we have $\tilde{\tau}_0=(\gamma^2+1)/2$ and $\delta(\gamma)\equiv(\tau_0-\tilde{\tau}_0)/\tau_0=0.052,\ 0.019,$

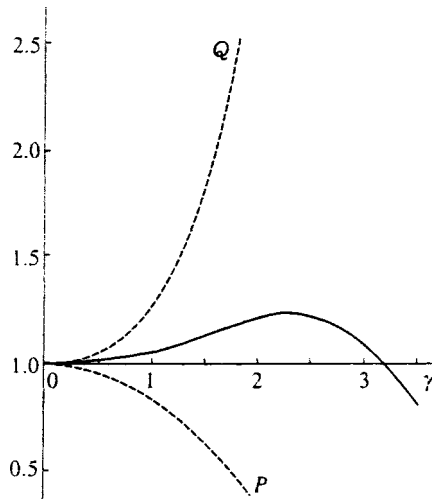

FIG. 4. The preexponential factor P, Coulomb factor Q (dashed curves), and their product (smooth curve) as functions of γ for $\theta=\pi/2$.

TABLE III. Accuracy of the adiabatic approximation.

γ	$g(\gamma,90°)$	δ_g, %	$P(\gamma,0)$	δ_P, %	$P(\gamma,90°)$	δ_P, %
2	1.15645	−1.33	0.5514	13.5	0.3847	−7.9
3	1.38876	−9.1(−3)	0.2995	0.25	5.406(−2)	−0.34
4	1.69336	−8.2(−6)	0.1466	0.034	2.233(−3)	−1.0(−3)
5	2.02800	−1.0(−9)	0.0674	4.5(−3)	3.13(−5)	−1.7(−7)
10	3.82538	–	9.08(−4)	2.0(−7)	3.29(−21)	–
15	5.67511	–	9.18(−6)	9.4(−12)	–	–

Note: The relative errors δ_g and δ_P are given in percents. A dash indicates that the corresponding value is less than 10^{-30}.

3.6(−4), and 1.2(−10), respectively, for $\gamma=1.5$, 2, 3, and 5. Expanding the function (1.8) in reciprocal powers of γ yields all of the terms in the series (B6).

Using the approximation (3.1), we find

$$g(\gamma,\theta)=\frac{3\tau_0}{2\gamma}\left\{1-\frac{1}{\gamma^2}\left[(\tau_0-1)\sin^2\theta+\frac{1}{3}\tau_0^2\cos^2\theta\right]\right\} \tag{3.3}$$

and

$$P(\gamma,\theta)=2\gamma^2(\tau_0\sin^2\theta+\tau_0^2\cos^2\theta)^{-1/2}e^{-\tau_0}, \tag{3.4}$$

as well as Eq. (2.6) for $Q(\gamma,\pi/2)$ (see Appendix C). Here one must substitute the approximation (1.8) for τ_0; for example, $\tau_0=\gamma$ for $\theta=0$, $\tau_0=\sqrt{2(1+\gamma^2)}-1$ for $\theta=45°$, and $\tau_0=(\gamma^2+1)/2$ for $\theta=90°$.

These asymptotic formulas are surprisingly accurate, starting with by $\gamma\sim3$ (see Table III, where δ denotes the relative error in the corresponding asymptote; thus, $\delta_g=(g-g_{as})/g$, where g is calculated using Eq. (2.1) and g_{as} is calculated using Eq. (3.3), while δ_P and δ_Q are defined analogously). Note that in the case of parallel fields, $g_{as}=Q_{as}\equiv1$, independently of γ, while $P_{as}=2\tau_0\exp(-\tau_0)$ and $\delta_P=\exp(-2\gamma)$. On the other hand, for $\theta=\pi/2$, we have

$$g_{as}=\frac{3\tau_0^2}{2(2\tau_0-1)^{3/2}},\quad P_{as}=\frac{2(2\tau_0-1)}{\sqrt{\tau_0}}e^{-\tau_0}. \tag{3.5}$$

with $\delta_g\approx-c_4\exp(-\gamma^2)$, $\delta_P\sim c_5\gamma^2\exp(-\gamma^2)$, and $c_4=2c_5=0.736$. This approximation is not so accurate for Q, because of the complicated form of the Coulomb integral (1.12). A comparison of the asymptote (2.6) with the numerical calculations shows that $\delta_Q=10.6\%$, 2.7%, and 0.69% for $\gamma=5$, 10, and 20, and δ_Q falls off as γ^{-2}, rather than exponentially.

Finally, using Eq. (3.1) it is easy to see how the action W accumulates during subbarrier motion. Let $\rho=r/b$, where $r\equiv(x^2+y^2+z^2)^{1/2}$ is the distance between the electron and nucleus (here the component $x(t)$ is purely imaginary, while y and z are real) and b is the barrier width (see Eq. (C5)). For $\theta<\pi/2$ we easily find

$$W(\rho)/W(1)=1-(1-\rho)^{3/2}+O(\gamma^{-1}). \tag{3.6}$$

It is interesting to note that the same dependence holds for $\gamma=0$, i.e., in the case of a purely electric field, when[3,36]

$$x=b_0\left[1-\left(\frac{t}{t_0}\right)^2\right],\quad W(t,t_0)=\frac{i\kappa^3}{3\mathscr{E}}\left[1-\left(\frac{t}{t_0}\right)^3\right] \tag{3.7}$$

($b_0=\kappa^2/2\mathscr{E}$). Thus, on going from $\gamma\ll1$ to $\gamma\gg1$, i.e., from a weak magnetic field to a strong one, the way the action varies along the subbarrier trajectory is essentially unchanged.

4. ASYMPTOTIC COEFFICIENT AT INFINITY

As Eq. (1.9) shows, the ionization probability for an atomic level is proportional to $|C_\kappa|^2$. This is not surprising, since for $\epsilon,h\ll1$ ionization takes place from the ''tail'' of the atomic wave function $\psi_0(r)$, while the barrier is wide ($\kappa b\sim1/\epsilon\gg1$). We now briefly discuss the properties of the coefficients C_κ.

For potentials with a Coulomb ''tail'' at infinity, $V(r)=-Z/r+o(r^{-2})$, the asymptote of the normalized radial wave function has the form

$$\chi_{\kappa l}(r)=2C_{\kappa l}\sqrt{\kappa}e^{-\kappa r}(\kappa r)^\eta\times\left[1+\frac{(l+\eta)(l+1-\eta)}{2\kappa r}+\ldots\right],\quad r\to\infty, \tag{4.1}$$

which determines the asymptotic coefficient $C_{\kappa l}$. Note that under the similarity transformation $\mathbf{r}\to\alpha\mathbf{r}$, in the Schrödinger equation

$$E_{\kappa l}\to\alpha^2E_{\kappa l},\quad \kappa\to\alpha\kappa,\quad V(\mathbf{r})\to\alpha^2V(\alpha\mathbf{r}),$$
$$\chi_l(r)\to\sqrt{\alpha}\chi_l(\alpha r),\quad C_{\kappa l}\to C_{\kappa l}. \tag{4.2}$$

Thus, the coefficients $C_{\kappa l}$ are dimensionless and scale-invariant, i.e., they depend only on the form of the potential V (and number of the level), but not on its depth and radius individually. For the ns-state of a hydrogenlike atom, we have

$$C_\kappa=2^{n-1}/n!,\quad l=0, \tag{4.3}$$

and these coefficients are independent of the nuclear charge. ($C_\kappa=1$ for $n=1,2$.)

The parameters κ and C_κ for neutral ($Z=1$) atoms and for negative ($Z=0$) and singly-charged positive ($Z=2$) ions are listed in Table I, which we have compiled from handbook data.[42] Here $C_\kappa=(2\kappa^{\eta+1/2})^{-1}A$, where A are the corresponding coefficients from Ref. 42. The values of C_κ for the ground states of atoms and ions lie within fairly narrow limits: for neutral atoms, from $C_\kappa=0.926$ for the Cs atom to $C_\kappa=1.22$ for Hg, and in most cases they are close to unity. (We have mostly chosen cases where the valence electron is in the s-state.) Thus, the asymptotic coefficients C_κ (and,

thereby, the ionization probability w) are insensitive to the form of the potential $V(r)$ outside the atomic core.

5. IMAGINARY TIME METHOD

In calculating the ionization probabilities, we used the imaginary time method (ITM). This made it possible, in particular, to find the Coulomb correction to the probability $w(\mathscr{E},\mathscr{H})$, which can hardly be found in any other way. Here we present some necessary explanations regarding the use of the ITM in ionization problems.

1) According to Feynman,[43] in order to calculate the probability of a transition from an initial into a final state, it is necessary to allow for contributions from all possible paths linking these states. In the case of subbarrier (tunneling) transitions and within the domain of applicability of the semiclassical approximation, it suffices to restrict attention to the class of smooth paths that satisfy the classical equations of motion (but with imaginary time t) and the natural boundary conditions: at $t=t_0$ the electron is bound to the atom and as $t\to\infty$ it escapes to infinity in specified external fields $\mathscr{E}$ and $\mathscr{H}$. The contribution of such a path to the tunneling probability is[36] $\exp(-2\,\mathrm{Im}\,W)$, where W is the reduced action,

$$W=\int_{t_0}^{0}\left[L+E_0-\frac{d}{dt}\,(\mathbf{p}\cdot\mathbf{r})\right]dt,\quad L=\frac{1}{2}\,m\mathbf{v}^2+\frac{e}{c}\,\mathbf{A}\cdot\mathbf{v}-e\varphi, \tag{5.1}$$

$E_0=-\kappa^2/2$ is the energy of the initial state, t_0 is the initial time for subbarrier motion, and $t=0$ is the time the particle escapes from under the barrier ($\ln W(t)$ no longer changes for $t>0$). In the following, we consider constant and uniform fields and a potential $U(r)$ with zero range (a δ-potential). Making the substitutions

$$\mathbf{A}=\frac{1}{2}\,[\mathscr{H}\mathbf{r}],\quad \varphi=-\mathscr{E}\mathbf{r}$$

in Eq. (5.1) and bearing the equation of motion and the initial condition $\mathbf{r}(t_0)=0$, in mind, we obtain

$$W=\frac{1}{2}\left\{e\mathscr{E}\int_{t_0}^{0}\mathbf{r}(t)dt+\kappa^2 t_0-m(\mathbf{r}\cdot\dot{\mathbf{r}})_{t=0}\right\}. \tag{5.2}$$

This then is the basic formula with which we shall calculate the ionization probability. We now proceed to some examples ($\hbar=e=m=1$).

2) In a purely electric field $\mathscr{E}=(0,0,\mathscr{E})$ the subbarrier trajectory of an electron can be characterized by the transverse momentum $\mathbf{p}_\perp$:

$$\mathbf{r}(t)=\left\{\mathbf{p}_\perp(t-t_0),\ \frac{1}{2}\,\mathscr{E}(t^2-t_0^2)\right\},$$
$$t_0=\frac{i}{\mathscr{E}}\sqrt{\kappa^2+p_\perp^2}=\frac{i\kappa}{\mathscr{E}}\left(1+\frac{p_\perp^2}{2\kappa^2}+\ldots\right), \tag{5.3}$$

which implies that $(\mathbf{r}\cdot\dot{\mathbf{r}})_{t=0}=-p_\perp^2 t_0$ and

$$W(p_\perp)=\frac{1}{6}\,\mathscr{E}^2 t_0^3+\frac{1}{2}\,(\kappa^2+p_\perp^2)t_0=\frac{i}{3\mathscr{E}}\,(\kappa^2+p_\perp^2)^{3/2}. \tag{5.4}$$

In the D-dimensional case, the total ionization probability for the s-level is given by

$$w_{sr}(\mathscr{E})=\frac{\kappa^2}{\pi}\,|C_\kappa|^2\int\exp\{-2\,\mathrm{Im}\,W(p_\perp)\}d^{D-1}\left(\frac{p_\perp}{\kappa}\right)$$
$$\approx\kappa^2\pi^{(D-3)/2}|C_\kappa|^2\epsilon^{(D-1)/2}e^{-2/3\epsilon},\quad \epsilon\to 0, \tag{5.5}$$

where C_κ is the dimensionless asymptotic coefficient of the (normalized) wave function in the short-range potential $U(r)$:

$$\psi_0(r)\approx\sqrt{\frac{\kappa}{\pi}}\,C_\kappa e^{-\kappa r}r^{-(D-1)/2},\quad r=\left(\sum_{i=1}^{D}x_i^2\right)^{1/2}\gg\kappa^{-1} \tag{5.6}$$

(in the case of a δ-potential, $\pi^{(D-3)/2}|C_\kappa|^2=1$, $\sqrt{\pi}/2$, and 1/2, respectively, for $D=1,2$, and 3). Here the distribution of escaping electrons with respect to $p_\perp$ is narrow (on an atomic scale): $\langle p_\perp\rangle\sim\sqrt{\epsilon\kappa}\ll\kappa$.

Equation (5.5) applies to ionization of an s-level bound by a short-range potential and reveals the dependence of the probability w_{sr} on the dimensionality of the space.[4)] Here we also give the ionization probability for the ground state of a D-dimensional ''hydrogen atom:''

$$w_{\mathrm{H}}(\mathscr{E})=\kappa^2\,\frac{2^{D-1}}{\Gamma((D+1)/2)}\,\epsilon^{-(D-1)/2}e^{-2/3\epsilon} \tag{5.7}$$

(for $\mathscr{E}=0$ the energy of the ground state is $E_0=-2(D-1)^{-2}$, so that $\kappa=2/(D-1)$ for $D>1$). Note that Eq. (5.7) follows directly from the general formula[2,3,6] for the ionization probability of the state $|n_1n_2m\rangle$ in a weak electric field if we note that the ground state in the D-dimensional Coulomb problem is equivalent to a nodeless ($n_1=n_2=0$, $n=\kappa^{-1}=(D-1)/2$) state of the three-dimensional hydrogen atom.[46] Here n_1, n_2, and m are the parabolic quantum numbers, with $n=n_1+n_2+|m|+1$.

A comparison of Eqs. (5.5) and (5.7) shows that they have the same exponential factors, but the preexponential factors are substantially different: for short-range potentials it decreases as the dimensionality D increases, while for the Coulomb potential it increases. This is explained by the fact that for $\epsilon\ll 1$, ionization takes place from the ''tail'' of the wave function $\psi_0(r)$, where the electron density

$$\psi_0^2(r)\propto r^{2\eta-(D-1)}e^{-2\kappa r},\quad \kappa r\gg 1 \tag{5.8}$$

is greater for larger values of the exponent $2\eta-(D-1)=(Z-1)(D-1)$.

3) If both $\mathscr{E}$ and $\mathscr{H}$ fields, are present, then the extremal[5)] path is no longer one-dimensional and has the form (1.5). The velocity of the particle along it is $\mathbf{v}=i\omega_c d\mathbf{r}/d\tau$, with

$$\mathbf{v}^2(\tau)=\left(\frac{e\mathscr{E}}{\omega_c}\right)^2\left\{\tau_0^2\left[\left(\coth\,\tau_0-\frac{1}{\tau_0}\right)^2+2\,\frac{\cosh\,\tau_0-\cosh\,\tau}{\tau_0\,\sinh\,\tau_0}\right]\sin^2\,\theta-(\tau_0^2\,\sin^2\,\theta+\tau^2\,\cos^2\,\theta)\right\}. \tag{5.9}$$

Equation (1.6) follows directly from the condition $\mathbf{v}^2(\tau_0)=-\kappa^2$, and for $\tau=0$ we find that the momentum of the particle as it escapes from under the barrier is directed perpendicular to the $(\mathscr{E},\mathscr{H})$ plane and has magnitude

$$p_0=v_x(0)=\frac{e\mathscr{E}}{\omega_c}\left(1-\frac{\tau_0}{\sinh\,\tau_0}\right)\sin\,\theta\approx\kappa\,\sin\,\theta\begin{cases}\frac{1}{6}\,\gamma, & \gamma\ll 1,\\ \gamma^{-1}, & \gamma\gg 1.\end{cases} \tag{5.10}$$

In contrast to the one-dimensional case, the point of escape from under the barrier is not a point at which the particle stops; however, the velocity components v_y and v_z, which are imaginary in the subbarrier motion, go to zero at this point.

It follows from Eq. (1.5) that for the extremal trajectory $(\mathbf{r}\cdot\dot{\mathbf{r}})_{t=0}=0$ and

$$W=-i\,\frac{\mathscr{E}^2}{2\omega_c^3}\,\tau_0\left[(\tau_0\,\coth\,\tau_0-1)\sin^2\,\theta+\frac{1}{3}\,\tau_0^2\,\cos^2\,\theta\right]+i\,\frac{\kappa^2}{2\omega_c}\,\tau_0=\frac{i}{3\epsilon}\,g(\gamma,\theta). \tag{5.11}$$

Using Eq. (1.6) and the equations $\mathscr{E}/\omega_c=\kappa/\gamma$, $\kappa^2/\omega_c=(\epsilon\gamma)^{-1}$, and $\mathscr{E}^2/\omega_c^3=(\epsilon\gamma^3)^{-1}$, we arrive at Eqs. (1.9) and (1.11).

4) To calculate the preexponential factor P, we have to include the contribution not only from the extremal trajectory, but from the entire bundle of ''classical'' trajectories close to it. We now illustrate this for parallel fields: $\mathbf{r}(t)=\{\rho(t),z(t)\}$, where

$$\boldsymbol{\rho}\equiv x+iy=i\,\frac{\mathbf{p}_\perp}{\omega_c}\,[\exp(-i\omega_c t)-\exp(-i\omega_c t_0)],$$

$$z=\frac{1}{2}\,\mathscr{E}(t^2-t_0^2) \tag{5.12}$$

(the fields $\mathscr{E}$ and $\mathscr{H}$ are directed along the z axis), with $\mathbf{r}(t_0)=0$ and $\dot{\mathbf{r}}(0)=\mathbf{p}_\perp$. The second condition of Eq. (1.4) takes the form

$$\mathscr{E}^2t_0^2+p_\perp^2\,\exp(-2i\omega_c t_0)=-\kappa^2$$

and determines the initial time for subbarrier motion,

$$t_0=\frac{i\kappa}{\mathscr{E}}\left(1+\frac{p_\perp^2}{2\kappa^2}\,e^{-2\tau_0}+\ldots\right),\quad p_\perp\ll\kappa.$$

The first two terms on the right-hand side of Eq. (5.2) yield a contribution

$$\frac{1}{6}\,\mathscr{E}^2t_0^3+\frac{1}{2}\,\kappa^2t_0=i\,\frac{\kappa^3}{3\mathscr{E}}\left[1+O\left(\left(\frac{p_\perp}{\kappa}\right)^4\right)\right],$$

and in the Gaussian ($\sim p_\perp^2$) approximation their dependence on the transverse momentum must be neglected. The last term of Eq. (5.2) remains:

$$(\mathbf{r}\cdot\dot{\mathbf{r}})_{t=0}=\frac{1}{2}\,(\dot{\boldsymbol{\rho}}\cdot\boldsymbol{\rho}^*+\dot{\boldsymbol{\rho}}^*\cdot\boldsymbol{\rho})_{t=0}=-p_\perp^2t_0\,\frac{\sin\,\omega_c t_0}{\omega_c t_0}.$$

Since $\omega_c t_0=i\gamma[1+O(p_\perp^2/\kappa^2)]$, we have, finally,

$$W(p_\perp)=\frac{i}{\epsilon}\left\{\frac{1}{3}+\frac{\sinh\,\gamma}{\gamma}\,\frac{p_\perp^2}{2\kappa^2}+\ldots\right\}. \tag{5.13}$$

Integrating over $d^2p_\perp$ yields a factor $\epsilon\gamma/\sinh\,\gamma$ in the preexponential factor (see Eqs. (1.9) and (1.17)). We now make some comments on the resulting solution.

a) In finding the subbarrier trajectories we allowed only for external $\mathscr{E}$ and $\mathscr{H}$ fields, which actually implies a zero-range approximation for the potential $U(r)$ binding the electron to the atom. Including a finite range of the forces, r_s, yields small corrections of order $r_s/b\sim\epsilon(\kappa r_s)$ in the preexponent, which we shall neglect.

b) The ITM provides an easy explanation for the reduction in ionization probability with increasing magnetic field. The barrier width b is given by

$$b\equiv\sqrt{\mathbf{r}^2(0)}=\frac{\kappa^2}{2\mathscr{E}}\,d(\gamma,\theta), \tag{5.14}$$

where

$$d(\gamma,\theta)=\left(\frac{\tau_0}{\gamma}\right)^2\sqrt{\cos^2\,\theta+\left[\frac{\tanh(\tau_0/2)}{\tau_0/2}\right]^2\sin^2\,\theta}=\begin{cases}1+\frac{\sin^2\,\theta}{36}\,\gamma^2+\ldots, & \gamma\ll 1,\\ \frac{1}{\cos\,\theta}-\frac{2\tan^2\,\theta}{\gamma}+\ldots, & \gamma\to\infty.\end{cases} \tag{5.14$'$}$$

For $\gamma\lesssim 1$ the barrier width is essentially constant and coincides with the width $b_0=\kappa^2/2\mathscr{E}$ for a constant electric field, with $\kappa b_0=1/2\epsilon\gg 1$. For $\gamma\gg 1$, for which the magnetic field exceeds the electric, however, $d(\gamma,\theta)$ increases; this widens the barrier and sharply reduces its permeability.

c) At the time of escape from under the barrier,

$$x=0,\quad\frac{y}{z}\equiv\tan\,\theta_0,\quad\theta_0=\arctan\left\{\frac{\tanh(\tau_0/2)}{\tau_0/2}\,\tan\,\theta\right\}.$$

When $\gamma\lesssim 1$ the particle moves along the electric field during subbarrier motion. If, on the other hand, $\gamma\gg 1$, then the subbarrier trajectory is ''clamped'' to the direction of the magnetic field ($\theta_0\ll\theta$).

5) Proceeding to the calculation of the Coulomb correction, we set $V=V_0+\delta V$, while the extremal subbarrier trajectory $\mathbf{r}_0(t)$ in the potential V_0 is assumed to be known (analytically or numerically), while the perturbation δV is small everywhere along it. The variation in the action is

$$\delta S=-\int_{t_0}^{0}\delta V(\mathbf{r}_0(t))\,dt. \tag{5.15}$$

In this case, $\delta V=-Z/r$, so the integral (5.15) diverges logarithmically as $r\to 0$ (or $t\to t_0$); in general, for the ground and low-lying excited states of atoms, the region $r\lesssim\kappa^{-1}$ cannot be treated semiclassically. This difficulty is overcome with the aid of a matching technique (cf. Ref. 1). Since $\epsilon\ll 1$, we have[6)] $b\gg r_0\gg\kappa^{-1}$, where b is the barrier width and $r_0=\sqrt{Z/\mathscr{E}}$ is the distance over which the Coulomb field of the atomic core (charge Z) is comparable to the external field $\mathscr{E}$. This makes it possible to introduce a matching point r_1 ($\kappa^{-1}\ll r_1\lesssim r_0$) such that for $r>r_1$ the Coulomb interaction distorts the subbarrier trajectory only weakly and Eq. (5.15) still works, while for $r<r_1$ the external fields can still be neglected, so that

$$\psi_0(r)\approx r^{-1}\exp\{-\kappa r+\eta\ln(\kappa r)+O(1)\}$$
$$=r^{-1}\exp\{-\mathrm{Im}[S_0(r)+\delta S(r)]\}.$$

Ultimately, for the Coulomb correction we obtain

$$\delta S_C=-i\eta\ln(\kappa r_1)+Z\int_{t_1}^{0}\frac{1}{\sqrt{\mathbf{r}_0^2(t)}}\,dt,$$

$$\exp(-2\,\mathrm{Im}\;\delta S_C)=\left(\frac{\epsilon}{2}\right)^{-2\eta}[Q(\gamma,\theta)]^{\eta},\qquad(5.16)$$

where $r_1=\sqrt{\mathbf{r}_0^2(t_1)}$ and $\eta=Z/\kappa$. This procedure is verified by the fact that the arbitrary matching point r_1 drops out of the final answer.

6) We illustrate Eq. (5.16) for an electric field, for which (see Eq. (5.3) for $p_\perp=0$) $r_0(t)=\mathscr{E}(t^2-t_0^2)/2$ and $t_0=i\kappa/\mathscr{E}$:

$$\delta S_C=-i\eta\ln(\kappa r_1)+\frac{2Z}{\mathscr{E}}\int_{t_1}^{0}\frac{dt}{t^2-t_0^2}$$
$$=-i\eta\ln\frac{(t_0+t_1)\kappa r_1}{t_0-t_1}.$$

Since $r_0(t)=i\kappa(t-t_0)+\ldots$ for $t\to t_0$, we have

$$\delta S_C=-i\eta\ln(-2i\kappa^2 t_0)=-i\eta\ln(2\kappa^3/\mathscr{E}),$$

$$\exp(-2\,\mathrm{Im}\;\delta S_C)=(2/\epsilon)^{2\eta},\qquad(5.17)$$

which coincides with the standard result obtained using the asymptotic solution of the Schrödinger equation in parabolic coordinates for the hydrogen atom[1–3] and for a short-range potential.[7,8] A comparison with these references shows how much simpler the calculation becomes when the ITM is used. Applying Eq. (5.16) to the subbarrier trajectory (1.5) leads to Eq. (1.12) for the Coulomb correction.

6. CONCLUSION

A semiclassical theory of the ionization of atoms and ions in electric and magnetic fields that allows for the Coulomb interaction between the electron and the atomic core, has been constructed. The resulting formulas are asymptotically exact for weak ($\epsilon,h\ll 1$) fields, and for $\mathscr{H}=0$ transform to the well known results[1–7] for the case of an electric field. Here the answer (Eq. (1.9)) contains only two constants (κ and C_κ), which characterize the given atomic level.

There are two questions which require further work.

1) We have discussed ionization of the s-level,[7)] but the ITM can be generalized to states with $l\neq 0$. Then, at short distances ($\kappa r\lesssim 1$), where the atomic potential is spherically symmetric, it is necessary to account for the removal of the degeneracy in the projection of m in the external fields $\mathscr{E}$ and $\mathscr{H}$. This can be done using Eqs. (A1)–(A5) from Appendix A, which solve for the correct wave functions in the zeroth approximation.[8)]

2) The ITM can be generalized to the relativistic case, which makes it possible to examine the ionization of levels whose binding energy is comparable to the rest energy mc^2. Allowing for corrections of order α^2, the ionization probability (to exponential accuracy) is

$$w(\mathscr{E},\mathscr{H})\propto\exp\left\{-\frac{2}{3\epsilon}(1-c_1\alpha^2\kappa^2)\right\},\qquad(6.1)$$

where

$$c_1=\frac{1}{30}\left[\frac{9}{4}-\left(\frac{\mathscr{H}}{\mathscr{E}}\right)^2\sin^2\theta\right],\quad \alpha=\frac{e^2}{\hbar c}=\frac{1}{137}.\qquad(6.2)$$

In particular, $c_1=3/40$ for a pure electric field and $c_1=1/24$ for crossed fields. In these cases the relativistic correction increases the ionization probability slightly, but it changes sign if $\mathscr{E}/\mathscr{H}<(2/3)\sin\theta$.

The relativistic version of the ITM can be used to obtain the analog of Eqs. (1.9) and (6.1) for a level with arbitrary energy (from mc^2 to $-mc^2$), but this question lies outside the scope of this article.[47]

We thank N. B. Delone and V. P. Kraĭnov for discussions of the results of this work, S. G. Pozdnyakov and A. V. Sergeev for help with the numerical calculations, and M. N. Markina for help in preparing the manuscript. This work was partially supported by the Russian Fund for Fundamental Research.

APPENDIX A

It has been shown[48] that in external static $\mathscr{E}$ and $\mathscr{H}$ fields, the energy spectrum of an electron bound by a short-range potential $U(r)$ of radius r_s is given by

$$\det[A_{lm}^{l'm'}(E)-B_l(E)\,\delta_{mm'}]=0,\qquad(A1)$$

which follows from the matching condition for the inner and outer wave functions in the overlap region,

$$r_s\ll r\ll\min\{\kappa^{-1},L_{\mathscr{E}},L_{\mathscr{H}}\}.$$

Here $\kappa=\sqrt{-2E}$, $L_{\mathscr{E}}=\mathscr{E}^{-1/3}$ is the electric length and $L_{\mathscr{H}}=(\mathscr{H}/c)^{-1/2}$ is the magnetic length.[10,48] Since $L_{\mathscr{E}}=\kappa^{-1}\epsilon^{-1/3}$ and $L_{\mathscr{H}}=\kappa^{-1}h^{-1/2}$ and assuming that conditions (1.2) are satisfied, in our case it is sufficient that

$$r_s\ll r\ll\kappa^{-1},\qquad(A2)$$

which can always be satisfied for $\kappa r_s\ll 1$, i.e, in the case of a shallow level.

The coefficients $A_{lm}^{l'm'}$ appear in the expansion of the solutions of the Schrödinger equation,

$$g_{lm}(\mathbf{r};E)=r^{-(l+1)}Y_{lm}\left(\frac{\mathbf{r}}{r}\right)+\ldots$$
$$+\sum_{l',m'}A_{lm}^{l'm'}(E)r^{l'}Y_{l'm'}\left(\frac{\mathbf{r}}{r}\right)+\ldots,\quad r\to 0, \tag{A3}$$

which satisfy the Sommerfeld radiation condition at infinity, and the $B_l(E)$ are determined by the scattering phases in the potential $U(r)$. In the effective radius approximation, we have

$$(2l+1)!!(2l-1)!!B_l(E)=k^{2l+1}\cot\delta_l^{(s)}(k)$$
$$=-\frac{1}{a_l}+r_lE, \tag{A4}$$

where a_l is the scattering length, r_l is the effective radius, and l is the orbital angular momentum. Solutions (A3) can be expressed in terms of the Green function G by

$$g_{lm}(\mathbf{r};E)=\frac{2\pi}{(2l-1)!!}Y_{lm}\left(\frac{\partial}{\partial\mathbf{r}'}\right)G(\mathbf{r},\mathbf{r}';E)|_{r'=0}. \tag{A5}$$

In particular, $g_{00}(\mathbf{r};E)=\sqrt{\pi}G(\mathbf{r},0;E)$. In turn, $G(\mathbf{r},\mathbf{r}';E)$ is given in terms of the time-dependent Green function by

$$G(\mathbf{r},\mathbf{r}';E)=\lim_{\delta\to+0}i\int_0^\infty G(\mathbf{r},t;\mathbf{r}',0)e^{i(E+i\delta)t}dt, \tag{A6}$$

which in uniform electric and magnetic fields can be factored[43,49]

$$G(\mathbf{r},t;\mathbf{r}',0)=\frac{1}{\sqrt{(2\pi it)^3}}\frac{u}{\sin u}\exp\{i[S_2(\boldsymbol{\rho},\boldsymbol{\rho}';t)$$
$$+S_1(z,z';t)]\}, \tag{A7}$$

where

$$S_1(z,z';t)=\frac{(z-z')^2}{2t}-\frac{1}{2}(z+z')F_\parallel t-\frac{1}{24}F_\parallel^2t^3, \tag{A8}$$

and

$$S_2(\boldsymbol{\rho},\boldsymbol{\rho}';t)=\frac{(\boldsymbol{\rho}-\boldsymbol{\rho}')^2}{2t}u\cot u+\frac{1}{2}\omega(x-x')(y+y')$$
$$-\frac{1}{2}[(x-x')(\cot u-u^{-1})+(y+y')]F_\perp t$$
$$-\frac{1}{8}F_\perp^2t^3\frac{1-u\cot u}{u^2}, \tag{A9}$$

$\mathbf{r}=\{\boldsymbol{\rho},z\}$, $F_\parallel=|e|\mathscr{E}\cos\theta$, and $F_\perp=|e|\mathscr{E}\sin\theta$ (for an electron $e<0$), we have introduced the variable

$$u=\omega t/2,\quad \omega=|\omega_c|, \tag{A10}$$

and chosen the gauge $\mathbf{A}=\{-\mathscr{H}y,0,0\}$. Here S_1 is the action for one-dimensional motion in a uniform field $F_\parallel$ and S_2 is the two-dimensional action for crossed fields $F_\perp$ and $\mathscr{H}$.

We write Eq. (A7) in the form

$$G(\mathbf{r},t;\mathbf{r}',0)=G_0(\mathbf{r},t;\mathbf{r}',0)+\xi(\mathbf{r},\mathbf{r}';t), \tag{A11}$$

where $G_0=(2\pi it)^{-3/2}\exp\{i(\mathbf{r}-\mathbf{r}')^2/2t\}$ is the free-particle Green function, while the function ξ does not lead to divergence as $\mathbf{r}\to\mathbf{r}'$ in Eq. (A6). In particular, as $t\to 0$,

$$\xi(0,0;t)=\frac{1}{24}\omega^2t^2-\frac{i}{24}F^2t^3+\ldots,\quad F^2=F_\parallel^2+F_\perp^2.$$

We first consider a δ-potential. $G(\mathbf{r},0;E)$ describes a quasistationary state of the particle in the combined field ($\mathscr{E}$, $\mathscr{H}$, and a well with zero range located at the point $\mathbf{r}'=0$) if the boundary condition

$$G(\mathbf{r},0;E)=\text{const}\cdot\{r^{-1}-\kappa_0+O(r)\},\quad r\to 0 \tag{A12}$$

is satisfied, where $\kappa_0=1/a_s$ is the coupling constant of the δ-potential, and $E=E_r-i\Gamma/2=-\kappa^2/2$ and $E_0=-\kappa_0^2/2$ are, respectively, the energy of the level in the external fields and the energy of the level for the free atom (ion). From Eqs. (A7)–(A9) and the integral

$$\int_0^\infty\exp\left\{i\left(Et+\frac{r^2}{2t}\right)\right\}\frac{dt}{t^{3/2}}=\frac{\sqrt{2\pi i}}{r}\exp(-\sqrt{-2E}r)$$

we obtain

$$G(\mathbf{r},0;E)=\frac{1}{2\pi i}\left\{\frac{1}{r}-\kappa+\frac{1}{\sqrt{2\pi i}}\right.$$
$$\left.\times\int_0^\infty\frac{dt}{t^{3/2}}e^{iEt}\xi(t)+\ldots\right\},\quad r\to 0. \tag{A13}$$

where

$$\xi(t)\equiv\xi(0,0;t)=\frac{u}{\sin u}\exp\left\{-\frac{i}{24}[F_\parallel^2+F_\perp^2f_1(u)]t^3\right\}-1,$$

$$f_1(u)=\frac{3}{u^2}(1-u\cot u)=1+\frac{1}{15}u^2+\ldots,\quad u\to 0, \tag{A14}$$

and the dots in Eq. (A13) represent terms that vanish as $r\to 0$. A comparison of Eq. (A13) with Eq. (A12) yields

$$\kappa-\kappa_0=\frac{1}{\sqrt{2\pi i}}\int_0^\infty\frac{dt}{t^{3/2}}e^{iEt}\xi(t), \tag{A15}$$

which can also be written in the same form as (1.14) with $u=\kappa^2ht/2$.

We now discuss some special cases. If $\mathscr{E}=0$, then $\Lambda\to 0$, $\Lambda\gamma^2=h^{-1}$, and, after some manipulations, Eq. (1.14) becomes the equation corresponding to a purely magnetic field (see Ref. 8 for $l=0$ and Ref. 48 for arbitrary angular momentum l).

When the magnetic field is turned off, we have

$$h\to 0,\quad u=0,\quad \xi(t)=\exp\left(-\frac{i}{24}\mathscr{E}^2t^3\right)-1.$$

Transforming to the variable $\tau=\kappa^2t/2$ in Eq. (A15), we obtain

$$\kappa_0=\kappa+\frac{\kappa}{\sqrt{4\pi i}}\int_0^\infty\frac{d\tau}{\tau^{3/2}}\,e^{-i\tau}\left[1-\exp\left(-\frac{i}{3}\,\epsilon^2\tau^3\right)\right],$$

$$\epsilon=\frac{\mathscr{E}}{\kappa^3}, \tag{A16}$$

which agrees with an earlier paper[50] for $l=0$ (here we take the opportunity to correct a misprint in that paper: $t^{-3/2}$ in the integral in Eq. (10) should read $t^{-1/2}$).

When the conditions (1.2) are satisfied, the integrals in Eqs. (1.14) and (A15) can be calculated by the method of steepest descent (the saddle point is $u_0=-i\tau_0$, where $\tau_0>0$ is the root of Eq. (1.6) given by the ITM). The calculations yield both the level shift

$$\Delta E=E_r-E_0=E_0\delta,\quad \delta=\frac{1}{2}\,|C_\kappa|^2\left(\epsilon^2-\frac{1}{3}\,h^2\right)+\ldots. \tag{A17}$$

and the level width $\Gamma=\hbar w$ (Eq. (1.9) with $\eta=0$ and the factor (1.13)). We emphasize that Eq. (A17) refers only to a weakly bound level in a short-range potential. For example, for the ground state in a Coulomb field,[1] $C_\kappa=1$ and $\delta=(9\epsilon^2-h^2)/2+\ldots$.

Proceeding in similar fashion, from Eqs. (A1)–(A5) we can obtain an equation for the energy $E=\kappa^2/2$, of the s-state including a correction for the effective radius of the system,

$$\kappa-\frac{1}{a_s}-\frac{1}{2}\,\kappa^2 r_s=\frac{1}{\sqrt{2\pi i}}\int_0^\infty\frac{dt}{t^{3/2}}\,e^{iEt}\left\{\frac{u}{\sin u}\times\exp\left[-\frac{1}{24}\,(F_\parallel^2+F_\perp^2 f_1(u))t^3\right]-1\right\}, \tag{A18}$$

which transforms to Eq. (A15) in the approximation of zero-range forces ($r_s=0$, $1/a_s=\kappa_0$). In particular, in a purely electric field,[50] we have

$$w(\mathscr{E})=\kappa^2C_\kappa^2\epsilon e^{-2/3\epsilon}[1-c_1\epsilon+O(\epsilon^2)], \tag{A19}$$

where

$$C_\kappa=\frac{1}{\sqrt{2}}\left(1+\frac{1}{2}\,\kappa_0 r_s+\ldots\right),\quad c_1=\frac{5}{3}\left(1+\frac{3}{20}\,\kappa_0 r_s+\ldots\right). \tag{A20}$$

Including the Coulomb interaction in the framework of the approach discussed here requires knowledge of the Green function, allowing for the effects of the external $\mathscr{E}$ and $\mathscr{H}$, as well as of the Coulomb field, which can scarcely be done in analytic form.

APPENDIX B

Here we introduce the expansion of the functions in Eq. (1.9). For $\gamma\ll 1$,

$$\tau_0(\gamma,\theta)=\sum_{k=0}^\infty c_k\gamma^{2k+1},\quad g(\gamma,\theta)=\sum_{k=0}^\infty\frac{3c_k}{2k+3}\,\gamma^{2k}, \tag{B1}$$

$$P(\gamma,\theta)=1-\frac{1}{6}\,\gamma^2+\frac{1}{3240}\,(20s^2-84s+63)\,\gamma^4+\ldots, \tag{B2}$$

and

$$Q(\gamma,\theta)=1+\frac{2}{9}\,s\gamma^2+\ldots, \tag{B3}$$

where $s=\sin^2\theta$ and

$$c_0=1,\quad c_1=\frac{s}{18},\quad c_2=\frac{1}{3240}\,(35s^2-24s),$$

$$c_3=\frac{11s^3}{3888}-\frac{s^2}{270}+\frac{s}{1050},\ldots \tag{B4}$$

The relation between the expansion coefficients for the functions τ_0 and g indicated in Eq. (B1) follows from the equation

$$\tau_0(\gamma,\theta)=\frac{1}{3\gamma}\,\frac{\partial}{\partial\gamma}\,[\gamma^3 g(\gamma,\theta)], \tag{B5}$$

whose validity can be verified by writing Eq. (1.11) in the form

$$\frac{1}{3}\,\gamma^3 g=\frac{1}{2}\left[\gamma^2\tau_0-\frac{1}{3}\,\tau_0^3-\left(\tau_0^2\coth\tau_0-\tau_0-\frac{1}{3}\,\tau_0^3\right)\sin^2\theta\right]$$

and then calculating $\partial(\gamma^3 g)/\partial\gamma-3\gamma\tau_0$.

For $\gamma\gg 1$ (strong magnetic field) and $\theta<\pi/2$, we obtain

$$\tau_0(\gamma,\theta)=\frac{\gamma}{\cos\theta}-\tan^2\theta\left(1-\frac{1}{2\gamma\cos\theta}+\ldots\right), \tag{B6}$$

$$g(\gamma,\theta)=\frac{1}{\cos\theta}-\frac{3\tan^2\theta}{2\gamma}\left(1-\frac{1}{\gamma\cos\theta}+\ldots\right), \tag{B7}$$

$$P=2\gamma\exp\left(-\frac{\gamma}{\cos\theta}+\tan^2\theta\right)[1+O(\gamma^{-1})], \tag{B8}$$

and

$$Q=\frac{1}{\cos^2\theta}\left[1+2\,\frac{\sin^2\theta}{\gamma}\ln\frac{\gamma}{\cos\theta}+O(\gamma^{-1})\right]. \tag{B9}$$

The case $\theta=\pi/2$ is special: Eq. (1.6) for $\tau_0(\gamma)$ takes the form $f(\tau_0)=\gamma^2$, where

$$f(\tau)=2\tau\coth\tau-\left(\frac{\tau}{\sin\tau}\right)^2-1=\sum_{k=1}^\infty\frac{2^{2k}(2k+1)}{(2k)!}\,B_{2k}\tau^{2k},$$

and B_{2k} are the Bernoulli numbers. Reverting this series and using Eq. (B5), we arrive at the expansions (2.3). If $\gamma\gg 1$, then Eq. (1.6) can be conveniently written in the form

$$\tau-2\sum_{n=1}^\infty(n\tau^2-\tau)e^{-2n\tau}=\frac{1}{2}\,(\gamma^2+1).$$

Solving iteratively, we arrive at the rapidly converging expansions

$$\tau_0(\gamma)=\sum_{n=0}^{\infty} P_n(z)e^{-nz}, \quad g(\gamma)=\gamma^{-3}\sum_{n=0}^{\infty} Q_n(z)e^{-nz}, \tag{B10}$$

where

$$z=\gamma^2+1, \quad P_0=\frac{1}{2}z, \quad P_1=\frac{1}{2}z^2-z,$$

$$P_2=-\frac{1}{2}z^4+3z^3-4z^2+z,$$

$$Q_0=-\frac{1}{2}\, Q_1=\frac{3}{8}z^2, \quad Q_2=\frac{3}{8}(z^4-4z^3+2z^2),$$

and for arbitrary n,

$$P_n(z)=\frac{2}{3}(Q_n'-nQ_n), \quad n=0,1,\ldots . \tag{B11}$$

Finally, for small angles $\theta\ll 1$, we have

$$\tau_0(\gamma,\theta)=\gamma(1+\tau_1\theta^2+\ldots), \quad g(\gamma,\theta)=1+g_1\theta^2+\ldots,$$

$$P=\frac{\gamma}{\sinh\,\gamma}(1-p_1\theta^2+\ldots), \quad Q=1+q_1\theta^2+\ldots\ , \tag{B12}$$

where

$$\tau_1(\gamma)=\frac{1}{2}(\coth\,\gamma-\gamma^{-1})^2,$$

$$g_1(\gamma)=\frac{3}{2\gamma^2}\left(1+\frac{1}{3}\,\gamma^2-\gamma\coth\,\gamma\right),$$

$$p_1(\gamma)=(\gamma\coth\,\gamma-1)\left(\frac{\gamma^2-1}{2\gamma^2}-\frac{\coth\,\gamma}{2\gamma}+\frac{1}{\sinh^2\,\gamma}\right). \tag{B13}$$

Note the extreme smallness of these coefficients for $\gamma\ll 1$:

$$\tau_1=\frac{1}{18}\,\gamma^2-\frac{1}{35}\,\gamma^4+\ldots, \quad g_1=\frac{1}{30}\,\gamma^2-\frac{1}{315}\,\gamma^4+\ldots,$$

$$p_1=\frac{7}{270}\,\gamma^4+\ldots\ . \tag{B14}$$

Thus, for weak magnetic fields, the dependence of the ionization probability on the angle between $\mathscr{E}$ and $\mathscr{H}$ can be neglected right up to $\theta\sim 1$ (in agreement with the numerical calculations of Fig. 2). In the other limit, we have

$$\tau_1=\frac{1}{2}-\gamma^{-1}+\ldots, \quad g_1=\frac{1}{2}-\frac{3}{2\gamma}+\ldots,$$

$$p_1=\frac{1}{2}\,\gamma-1+O(\gamma^{-2}), \quad \gamma\to\infty \tag{B15}$$

(all these coefficients increase monotonically with increasing γ).

APPENDIX C

We now discuss the details of some of the calculations in Sec. 3. For the components of the velocity of subbarrier motion, Eq. (3.1) gives ($e=-1$)

$$v_x=\dot{x}=\frac{\kappa}{\gamma}(1-\tau_0\, e^{\tau-\tau_0})\sin\,\theta, \quad \dot{y}=-i\,\frac{\kappa}{\gamma}\,\tau_0\, e^{\tau-\tau_0}\sin\,\theta,$$

$$\dot{z}=-i\,\frac{\kappa}{\gamma}\,\tau\cos\,\theta, \tag{C1}$$

which implies Eq. (3.2). Furthermore,

$$W(t_0,t)=\int_{t_0}^{t}(L+E_0)dt=\frac{i}{\omega_c}\left\{\frac{1}{2}\,\kappa^2(\tau_0-\tau)-\int_{\tau}^{\tau_0}L\,d\tau\right\}, \tag{C2}$$

$$L=\frac{1}{2}\,\dot{\mathbf{r}}^2-\frac{1}{2}\,\omega_c(x\dot{y}-y\dot{x})-\mathscr{E}(y\,\sin\,\theta+z\,\cos\,\theta)$$

$$=\frac{\kappa^2}{2\gamma^2}\{(\tau_0^2-2\tau^2)\cos^2\,\theta+[\tau_0+1$$

$$+\tau_0(\tau_0-\tau-3)e^{\tau-\tau_0}]\sin^2\,\theta+\ldots\}.$$

An elementary integration yields

$$W(t_0,t)=\frac{i\,K}{2}\left\{\tau_0-\tau-\frac{1}{\gamma^2}\left[\frac{1}{3}(\tau_0^3-3\tau_0^2\tau+2\tau^3)\right.\right.$$

$$\left.\left.\times\cos^2\,\theta+(\tau_0^2-\tau_0-(\tau_0+1)\tau)\sin^2\,\theta\right]\right\}, \tag{C3}$$

where $K=\kappa^2/\omega_c=1/\epsilon\gamma$, $\tau_0-\tau\gg 1$, and we have discarded terms proportional to $\exp(\tau-\tau_0)$. Upon escape from under the barrier, Eq. (C3) transforms to Eq. (3.3) for $g(\gamma,\theta)$. From Eq. (3.1) we also find

$$r=(\mathscr{E}/\omega_c^2)\xi(\tau),$$

$$\xi=\tau_0\left\{\frac{1}{4}\,\tau_0^2(1-u^2)^2\,\cos^2\,\theta+(1-u)\right.$$

$$\left.\times(1-u-2e^{-\tau_0(1-u)})\sin^2\,\theta\right\}^{1/2}, \tag{C4}$$

where $u=\tau/\tau_0$. Here the barrier width is

$$b=\frac{\mathscr{E}}{\omega_c^2}\,\xi(0)=\frac{\kappa^2}{2\mathscr{E}}\,\frac{\tau_0}{\gamma^2}\,\sqrt{\tau_0^2\,\cos^2\,\theta+4\sin^2\,\theta}, \tag{C5}$$

and in the limit $\gamma\to\infty$, we have

$$b=\frac{\kappa^2}{2\mathscr{E}}$$

$$\times\begin{cases}\dfrac{1}{\cos\,\theta}-2\gamma^{-1}\tan^2\,\theta+\ldots, & \gamma\gg\dfrac{1}{\cos\,\theta}\ (0\leqslant\theta<\pi/2),\\ 1+\gamma^{-2}+\ldots, & \theta=\pi/2.\end{cases} \tag{C6}$$

We calculate the Coulomb correction using Eq. (1.12). If $\theta < \pi/2$, then $\xi/\gamma = \tau_0(1-u^2)/2$, the integral in Eq. (1.12) is given by $J = \ln 2$, and $Q(\infty) \approx (\tau_0/\gamma)^2 = 1/\cos^2\theta$. In the special case $\theta = \pi/2$, we have

$$\xi = \tau_0 \sqrt{y[2(1-e^{-\tau_0 y}) - y]}, \quad y = 1-u, \tag{C7}$$

and $\tau_0 = (\gamma^2+1)/2$. Introducing the matching point a such that $1/\tau_0 \ll a \ll 1$, we separate the integral J into two parts and proceed to Eq. (2.6), where the constant is given by

$$c_3 = \frac{1}{4}\exp\{2(J_1+J_2-2)\}, \tag{C8}$$

where

$$J_1 = \int_0^1 \left(\frac{1}{\sqrt{x(1-e^{-x})}} - \frac{1}{x} \right) dx,$$

$$J_2 = \int_1^\infty \left(\frac{1}{\sqrt{1-e^{-x}}} - 1 \right) \frac{dx}{\sqrt{x}}.$$

Numerically, we have $J_1 = 0.25433$ and $J_2 = 0.16692$, whereupon $c_3 = 0.01063$. Numerical calculations of $Q(\gamma)$ confirm the validity of the asymptote (2.6), which, however, holds only for $\gamma \gtrsim 15$.

*)E-mail: karnak@theor.mephi.msk.su

1)Here we have in mind analytic formulas, and not numerical calculations.[33] We note, however, that some partial results have, in fact, been obtained elsewhere.[34,35]

2)Strictly speaking, these conditions correspond to a zero range of action for the forces binding an electron to the atomic core. It can, however, be shown that allowance for the finite radius (range) of the forces yields corrections of order $\epsilon \ll 1$ in the multiplicative factor preceding the exponential, which we shall neglect (see Eq. (A19), in particular).

3)These expressions were obtained by analytic continuation of the standard formulas for the motion of a charged particle in constant and uniform fields.[41]

4)The transition from the physical value $D=3$ to arbitrary D has been developed extensively in recent years and is widely used in theoretical physics (the $1/D$-expansion, or dimensional scaling); see Refs. 44 and 45, and references therein.

5)That is, the subbarrier path that minimizes Im W and determines the most probable path for tunneling of the particle. It is selected by the boundary conditions (1.4).

6)Note that $r_0/b \sim \sqrt{Z\mathscr{E}}\kappa^{-2} \sim \sqrt{ZF}$, where $F = n^4\mathscr{E}/\mathscr{E}_a$ is the reduced electric field, which is conveniently introduced in the analysis of the Stark effect for Rydberg states.[21–23] The barrier in the potential $U_2(\eta)$ vanishes for a ''critical'' field $F = F_*$, where F_* varies between 0.130 for a state of the type $|0,n-1,0\rangle$ and 0.383 for states $|n-1,0,0\rangle$ with $n \gg 1$. The condition for applicability of the equations of perturbation theory for the Coulomb interaction (5.16) is $F \ll F_*$.

7)This case is encountered most often. Thus, among the neutral atoms (from hydrogen to uranium) the valence electron is in an s-state in 61 cases, in a p-state in 30 cases, and in an $l=2$ state for only one atom (Pd).[42]

8)This method has been used[48] to calculate the quasienergy spectrum of the p-levels of an electron in a short-range potential and the field of a circularly polarized wave. Here it is important to include degeneracy with respect to m, but it only changes the partial widths Γ_{lm}, while the total width $\Gamma_l = \Sigma_{m=-l}^{l} \Gamma_{lm}$ is the same as that calculated previously.[3,4]

[1] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics. Nonrelativistic Theory*, Pergamon, New York (1974).
[2] B. M. Smirnov and M. I. Chibisov, Zh. Éksp. Teor. Fiz. **49**, 841 (1965) [Sov. Phys. JETP **22**, 585 (1965)].
[3] A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **50**, 1393 (1966) [Sov. Phys. JETP **23**, 924 (1966)]; Zh. Éksp. Teor. Fiz. **51**, 309 (1966) [Sov. Phys. JETP **24**, 207 (1966)].
[4] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **50**, 255 (1966) [Sov. Phys. JETP **23**, 168 (1966)].
[5] J. D. Bekenstein and J. B. Krieger, Phys. Rev. **188**, 130 (1969).
[6] T. Yamabe, A. Tachibana, and H. J. Silverstone, Phys. Rev. A **16**, 877 (1977).
[7] Yu. N. Demkov and G. F. Drukarev, Zh. Éksp. Teor. Fiz. **47**, 918 (1964) [Sov. Phys. JETP **20**, 614 (1964)].
[8] Yu. N. Demkov and V. N. Ostrovskiĭ, *Zero-Range Potentials and their Applications in Atomic Physics*, Plenum Press, New York (1988).
[9] V. S. Lisitsa, Usp. Fiz. Nauk **153**, 379 (1987) [Sov. Phys. Usp. **30**, 927 (1987)].
[10] L. P. Gor'kov and I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **53**, 717 (1967) [Sov. Phys. JETP **26**, 449 (1967)].
[11] B. R. Johnson, K. F. Scheibner, and D. Ferrely, Phys. Rev. Lett. **51**, 2280 (1983).
[12] Yu. N. Demkov, B. S. Monozon, and V. N. Ostrovskiĭ, Zh. Éksp. Teor. Fiz. **57**, 1431 (1969) [Sov. Phys. JETP **30**, 775 (1969)].
[13] E. A. Solov'ev, Zh. Éksp. Teor. Fiz. **85**, 109 (1983) [Sov. Phys. JETP **58**, 63 (1983)].
[14] N. L. Manakov, S. I. Marmo, and V. D. Ovsyanikov, Zh. Éksp. Teor. Fiz. **91**, 404 (1986) [Sov. Phys. JETP **64**, 236 (1986)].
[15] V. M. Vaĭberg, V. S. Popov, and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **98**, 847 (1990) [Sov. Phys. JETP **71**, 470 (1990)].
[16] V. S. Melezhik, Phys. Rev. A **48**, 4528 (1993).
[17] R. J. Damberg and V. V. Kolosov, J. Phys. B **9**, 3149 (1976); **11**, 1921 (1978).
[18] H. J. Silverstone, Phys. Rev. A **18**, 1853 (1978); Phys. Rev. Lett. **43**, 1498 (1979).
[19] S. P. Alliluev, V. L. Eletsky, and V. S. Popov, Phys. Lett. A **73**, 103 (1979); S. P. Alliluev, V. M. Vaĭnberg, V. L. Eletskiĭ, and V. S. Popov, Zh. Éksp. Teor. Fiz. **82**, 77 (1982) [Sov. Phys. JETP **55**, 46 (1982)].
[20] V. Franceschini, V. Grecchi, and H. J. Silverstone, Phys. Rev. A **32**, 1338 (1985).
[21] V. M. Vaĭnberg, V. D. Mur, V. S. Popov *et al.*, JETP Lett. **44**, 9 (1986); JETP Lett. **46**, 225 (1987); Zh. Éksp. Teor. Fiz. **93**, 450 (1987) [Sov. Phys. JETP **66**, 258 (1987)].
[22] V. S. Popov, V. D. Mur, A. V. Sergeev *et al.*, Phys. Lett. A **124**, 77 (1987); **149**, 418, 425 (1990).
[23] V. S. Popov, Phys. Lett. A **173**, 63 (1993).
[24] F. M. Fernandez, Phys. Rev. A **54**, 1206 (1996).
[25] M. C. Chu and H. Friedrich, Phys. Rev. A **28**, 1423, 3651 (1983); **29**, 675 (1984).
[26] J. Main and G. Wunner, J. Phys. B **27**, 2835 (1994).
[27] J. H. Wang and C. S. Hsue, Phys. Rev. A **52**, 4508 (1995).
[28] I. Seipp and W. Shweizer, Astron. Astrophys. **318**, 990 (1997).
[29] L. P. Kotova, A. M. Perelomov, and V. S. Popov, Zh. Éksp. Teor. Fiz. **54**, 1151 (1968) [Sov. Phys. JETP **27**, 616 (1968)].
[30] G. F. Drukarev and B. S. Monozon, Zh. Éksp. Teor. Fiz. **61**, 956 (1971) [Sov. Phys. JETP **34**, 509 (1971)].
[31] S. P. Andreev and V. A. Polunin, JETP Lett. **42**, 190 (1985).
[32] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1964)].
[33] B. A. Zon, N. L. Manakov, and L. P. Rapoport, Zh. Éksp. Teor. Fiz. **61**, 968 (1971) [Sov. Phys. JETP **34**, 515 (1971)]; L. P. Rapoport, B. A. Zon, and N. L. Manakov, *The Theory of Many-Photon Processes in Atoms* [in Russian], Atomizdat, Moscow (1978).
[34] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **52**, 223 (1967) [Sov. Phys. JETP **25**, 145 (1967)].
[35] A. M. Perelomov and V. S. Popov, Zh. Éksp. Teor. Fiz. **52**, 514 (1967) [Sov. Phys. JETP **25**, 336 (1967)].
[36] V. S. Popov, V. P. Kuznetsov, and A. M. Perelomov, Zh. Éksp. Teor. Fiz. **53**, 331 (1967) [Sov. Phys. JETP **26**, 222 (1967)].
[37] B. M. Karnakov, V. D. Mur, and V. S. Popov, JETP Lett. **65**, 405 (1997).
[38] V. S. Popov and A. V. Sergeev, JETP Lett. **63**, 417 (1996).
[39] V. S. Popov, B. M. Karnakov, and V. D. Mur, Phys. Lett. A **229**, 306 (1997).
[40] A. S. Davydov, *Solid State Theory* [in Russian], Nauka, Moscow (1988).

[41] L. D. Landau and E. M. Lifshitz, *Classical Theory of Fields* [in Russian], Nauka, Moscow (1988).
[42] A. A. Radtsig and B. M. Smirnov, *Parameters of Atoms and Atomic Ions* [in Russian], Énergoatomizdat, Moscow (1968).
[43] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York (1965).
[44] D. R. Herschbach, J. Avery, and O. Goscinsky (eds), *Dimensional Scaling in Chemical Physics*, Kluwer Academic, Dordrecht (1993).
[45] C. A. Tsipis, V. S. Popov, D. R. Herschbach, and J. S. Avery (eds), *New Methods in Quantum Theory*, Kluwer Academic, Dordrecht (1996).
[46] S. P. Alliluev and V. S. Popov, Zh. Éksp. Teor. Fiz. **104**, 3569 (1993) [JETP **77**, 701 (1993)].
[47] V. S. Popov, V. D. Mur, and B. M. Karnakov, JETP Lett. **66**, 229 (1997).
[48] S. P. Andreev, B. M. Karnakov, and V. D. Mur, JETP Lett. **37**, 187 (1983); Teor. Mat. Fiz. **64**, 287 (1985).
[49] L. I. Magarill and S. K. Savvinykh, Zh. Éksp. Teor. Fiz. **60**, 175 (1971) [Sov. Phys. JETP **33**, 97 (1971)].
[50] S. P. Andreev, B. M. Karnakov, V. D. Mur, and V. A. Polunin, Zh. Éksp. Teor. Fiz. **86**, 866 (1984) [Sov. Phys. JETP **59**, 506 (1984)].

Translated by D. H. McNeill

Dynamic theory of two-photon correlators in the spectroscopy of single impurity centers

I. S. Osad'ko*)

Moscow State Pedagogical University, 119882 Moscow, Russia
(Submitted 17 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1606–1631 (May 1998)

A dynamic theory of two-photon correlators measured in experiments with single molecular impurity centers has been developed. The theory takes into account the interaction between optically active electrons of an impurity center and phonons, excitations in tunneling systems of polymers and glasses, and transverse electromagnetic field. Both the correlator measured in the start-stop regime and the ''full'' correlator have been analyzed, and equations for these correlators have been derived. An equation relating these two correlators has been also obtained. The effect of the triplet level of an impurity molecule on the correlators leading to bunching of spontaneously emitted photons has been studied. The two-photon correlators have been calculated numerically, and their dependence on the incident light frequency and time between the pairs of detected photons in various realistic situations has been derived.
 [S1063-7761(98)00505-8]

1. INTRODUCTION

The basic measurements in spectroscopy of ensembles of impurity centers are of absorption and fluorescence bands. In most cases, these bands are notably broader than bands of a single impurity center owing to inhomogeneous broadening. In real crystals, the local environment of each impurity center always varies slightly, which causes some inhomogeneous broadening of the spectral lines due to impurities in solid solutions. This inhomogeneous broadening can be two or three orders of magnitude larger and reach several hundreds of wave numbers if a polymer or glass is doped by impurity molecules.

The principal problem of the selective spectroscopy of impurity centers is getting round this inhomogeneous broadening, which obscures the fine features of impurity bands.[1] The techniques of persistent spectral hole burning[2] and fluorescence line narrowing[3] have allowed researchers to get rid of most of the inhomogeneous broadening, but not all. The point is that these two techniques select some molecules from an ensemble using the frequency of their phononless line. Leaving aside the impossibility of absolutely accurate frequency selection, selected molecules always have an uncertainty in the amplitude of electron–phonon interaction or interaction with tunneling systems of polymers and glasses in the local impurity environment. This means that even the spectra of molecular subensembles selected using resonant excitation are different from those of individual molecules, and the problem of measuring a really uniform spectrum can be resolved only on the level of individual molecules.

The problem of measuring optical spectra of individual impurity centers was solved several years ago,[4,5] and this event initiated intense development of single-molecule spectroscopy.[6,7] At the present time, the following features of individual molecules have been measured: absorption line shape,[8] vibronic spectra,[9] magnetic resonance of triplet levels,[10,11] static Stark effect,[12] dynamic Stark effect,[13] and distribution over optical line widths of individual impurity centers.[14]

Optical parameters of individual impurity centers are usually measured at liquid-helium temperatures under cw optical excitation by a laser with a spectral line width of several megaherz. Photons sequentially emitted by a center are detected. Measurements are performed in the photon counting regime, and the measured parameter is the correlator in pairs of photons emitted by the studied impurity center. For this reason, development of the theory of two-photon correlators is, undoubtedly, a key problem in single-molecule spectroscopy. This paper is devoted to the solution of this problem.

The paper is organized as follows. Section 2 describes specific features of measurements in the start-stop regime and gives an expression for a two-photon correlator measured in the start-stop regime. Section 3 presents derivation of the so-called full two-photon correlator. In these two sections the interaction between molecules and phonons is neglected. In Sec. 4 we will consider the effects of phonons and tunneling systems of polymers and glasses on two-photon correlators. Section 5 gives equations relating correlators of these two types. In Sec. 6 we will discuss the correlators as functions of time and pumping laser light frequency. Finally, Sec. 7 considers the effect of a triplet level on the full two-photon correlator, which shows itself in the grouping of emitted photons on the time axis and is termed ''photon bunching.''

2. TWO-PHOTON CORRELATOR MEASURED IN THE START-STOP REGIME

Consider a situation when a single atom is exposed to cw monochromatic laser radiation. In each elementary event, the atom absorbs and after some time emits one photon. A set of sequentially emitted photons flies through space with random

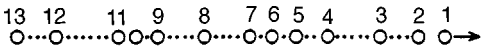

FIG. 1. Sequence of photons emitted by an atom exposed to cw light. The time intervals between photons in pairs (2, 3), (4, 6), and (12, 13) are equal.

time intervals between individual photons, as shown in Fig. 1. When fed into a PMT cathode, these photons generate a sequence of electric pulses apparently randomly distributed in time. There is, however, a certain regularity in the times at which photons are detected by the PMT. This can be checked by setting two PMTs on the beam line and detecting two photons following one another after a certain time interval t_0. This kind of measurement is called the start-stop regime. For example, if the interval between photons 2 and 3 in Fig. 1 is t_0, the second event will be detected only after the arrival of photons 12 and 13.

If the total time of the experiment on detection of certain photon pairs is relatively short, the number of such events should greatly fluctuate from one experiment to another. The longer the measurement time, the smaller the ratio of the fluctuation amplitude to the average number of interesting events. Therefore, after a sufficiently long measurement time, the ratio

$$\frac{N_2(t_0,t)}{N_2(t)}=S(t_0), \tag{1}$$

where the numerator is the number of interesting events detected during the time interval t, and the denominator is the total number of pairs arriving at the PMT during the same time, should be a function only of t_0 and independent of t.

Consider the probabilities W_0, W_1, and $W_{\mathbf{k}}$ of detecting the atom in the ground state, excited state, and after emission of a spontaneous photon with momentum $\mathbf{k}$, respectively. Let us analyze the relation between the ratio (1), which is, of course, less than unity, and these three probabilities. The time when the PMT detects the first emitted photon should be defined as $t=0$. At this moment we have $W_0=1$, i.e., the atom is in the ground state, and the other probabilities are zero. If at time t_0 the second PMT detects a photon spontaneously emitted by the atom permanently exposed to the laser field, the probability of this event is the sum of all the probabilities $W_{\mathbf{k}}$ and, obviously, equals the ratio (1), i.e.,

$$\sum_{\mathbf{k}} W_{\mathbf{k}}(t_0)=S(t_0). \tag{2}$$

Usually experimenters measure not the total number of interesting events, but their number per unit time, i.e., the count rate. This count rate is determined by the following expression:

$$s(t_0)=\dot{S}(t_0)=\frac{d}{dt_0}\sum_{\mathbf{k}} W_{\mathbf{k}}(t_0). \tag{3}$$

Now let us calculate the probability.

Let H be the Hamiltonian of a closed system, for example, an atom plus an electromagnetic field, and $|m\rangle$ the wave functions of the system state at the initial moment. This state changes with time in accordance with the Schrödinger equation, whose solution is

$$|m,t\rangle=\exp(-itH/\hbar)|m\rangle. \tag{4}$$

The final state $|m,t\rangle$ in the general case is a superposition of many atomic states, including, for example, $|l\rangle$. Then the integral $\langle l|m,t\rangle$ is the probability amplitude of detecting the system in state $|l\rangle$, although it was in state $|m\rangle$ at the initial time. Let us denote this amplitude as

$$G_{lm}(t)=-i\langle l|m,t\rangle=-i\langle l|\exp(-itH/\hbar))|m\rangle. \tag{5}$$

Here the factor $-i$ is introduced for convenience in calculating Laplacian components of amplitudes. It does not affect the results because the probability measured in experiments is $|G_{lm}(t)|^2$.

By expressing the Hamiltonian in the form of the sum $H=H_0+\Lambda$, where Λ is responsible for the interaction between the atom and the transverse electromagnetic field, differentiating both sides of Eq. (5) with respect to time, and using the completeness condition for the set of wave functions, we obtain an equation for the amplitudes:

$$\dot{G}_{lm}(t)=-i\omega_{lm}G_{lm}(t)-\frac{i}{\hbar}\sum_s \Lambda_{ls}G_{sm}(t), \tag{6}$$

where the frequency is expressed in terms of the eigenvalues of the Hamiltonian H_0 as $\omega_{lm}=(E_l-E_m)/\hbar$.

Let the initial state $|m\rangle$ be characterized by the following wave function and energy:

$$|m\rangle=|0\rangle|n\rangle\equiv\left|\begin{matrix}0\\ n\end{matrix}\right\rangle, \quad E_m\equiv E_n=\hbar\omega_0 n+E_0, \tag{7}$$

where E_0 is the energy of zero-field oscillations and n is the number of photons in the laser mode interacting with the atom. Initially the atom is in the state whose energy has been set to zero. This initial state of the ''atom + field'' system is not steady. Under the action of interaction operator Λ, the atom should absorb a photon of the laser mode, and the system gets to the state $|1\rangle|n-1\rangle$ with energy $E+\hbar\omega_0(n-1)+E_0$, where E is the atom excitation energy, etc. The chain of states connected by operator Λ is as follows:

$$\left|\begin{matrix}0\\ n\end{matrix}\right\rangle\leftrightarrow\left|\begin{matrix}1\\ n-1\end{matrix}\right\rangle\leftrightarrow\left|\begin{matrix}0\\ n-1,\mathbf{k}\end{matrix}\right\rangle\leftrightarrow\left|\begin{matrix}1\\ n-2,\mathbf{k}\end{matrix}\right\rangle\leftrightarrow\ldots \tag{8}$$

Hereafter we use the resonant approximation, i.e., we take into account only the part of Λ which does not change the full number of excitations.

By applying the quantum states (8) to Eq. (6) and omitting the subscript of the initial state of amplitudes, we obtain the following infinite chain of coupled equations:

$$\begin{aligned}
&\dot{G}^0_n(t)=-i\Lambda^* G^1_{n-1},\\
&\dot{G}^1_{n-1}(t)=-i\Delta G^1_{n-1}(t)-i\left[\Lambda G^0_n(t)+\sum_{\mathbf{k}}\lambda_{\mathbf{k}}G^0_{n-1\mathbf{k}}(t)\right],\\
&\dot{G}^0_{n-1\mathbf{k}}(t)=-i\Delta_{\mathbf{k}}G^0_{n-1\mathbf{k}}(t)-i\lambda^*_{\mathbf{k}}G^1_{n-1}(t)+\ldots,
\end{aligned} \tag{9}$$

Here $\Delta=\Omega-\omega_0$, $\Delta_{\mathbf{k}}=\omega_{\mathbf{k}}-\omega_0$, Ω is the electron excitation frequency, $\omega_{\mathbf{k}}$ is the frequency of the spontaneously emitted photon, ω_0 is the laser photon frequency, and the matrix elements of the operator Λ are determined by the following equations:

$$\Lambda=\frac{\Lambda_{10}}{\hbar}=-i\Omega\sqrt{\frac{4\pi}{\hbar\omega_0}}\mathbf{d}\cdot\mathbf{e}_0\sqrt{\frac{n}{V}}=-i\frac{\mathbf{d}\cdot\mathbf{E}}{\hbar}=-i\chi,$$
$$\lambda_{\mathbf{k}}=-i\Omega\sqrt{\frac{4\pi}{\hbar\omega_{\mathbf{k}}}}\,\mathbf{d}\cdot\mathbf{e}_{\mathbf{k}}\sqrt{\frac{1}{V}}, \tag{10}$$

where V is the volume occupied by the laser field, $\mathbf{e}$ and n are the polarization vector of photons and their number, $\mathbf{d}$ is the dipole moment of the atomic transition, and $\mathbf{E}$ and χ are the electric field vector of the laser wave and the so-called Rabi frequency.

The first three lines of Eq. (9) contain the amplitudes needed for our analysis of the start-stop correlator. In calculating these amplitudes, we take into account only the terms given in Eq. (9) in the explicit form and neglect the rest. A fairly simple and persuasive argument in favor of this approximation is that it will allow us to derive optical Bloch equations. A more rigorous argumentation is given in the Appendix.

Using Eq. (9) and the complex-conjugate equations, and the formula

$$\frac{d}{dt}|G(t)|^2=\dot{G}G^*+G\dot{G}^*, \tag{11}$$

we easily obtain the following conservation law:

$$\frac{d}{dt}\left(|G_n^0(t)|^2+|G_{n-1}^1(t)|^2+\sum_{\mathbf{k}}|G_{n-1\mathbf{k}}^0(t)|^2\right)=0. \tag{12}$$

This is the law of probability conservation in the start-stop regime, which proves that our approximation has not broken the unitarity of the approximate amplitudes. Specifically, given that these amplitudes are related to the previously introduced probabilities by the formulas $|G_n^0(t)|^2=W_0$, $|G_{n-1}^1(t)|^2=W_1$, and $|G_{n-1\mathbf{k}}^0(t)|^2=W_{\mathbf{k}}$, and that $W_0(0)=1$, with due account of Eq. (12) we obtain

$$W_0(t)+W_1(t)+\sum_{\mathbf{k}}W_{\mathbf{k}}(t)=1. \tag{13}$$

These probabilities are, by definition, diagonal elements of the density matrix. Using Eqs. (11) and (9), we can derive similar equations for nondiagonal elements of the density matrix. By using the formula

$$\sum_{\mathbf{k}}\lambda_{\mathbf{k}}G_{n-1\mathbf{k}}^0(t)=-\frac{i}{2T_1}G_{n-1}^1(t), \tag{14}$$

whose validity was discussed previously[15] and where

$$\frac{1}{T_1}=\frac{4d^2}{3\hbar}\left(\frac{\Omega}{c}\right)^3 \tag{15}$$

is the rate of spontaneous emission from the excited atom, we can exclude from the three upper lines of Eq. (9) the amplitudes that are the functions of the wave vector $\mathbf{k}$ of the spontaneously emitted photon and derive the two equations

$$\dot{G}_n^0(t)=-i\Lambda^*G_{n-1}^1,$$
$$\dot{G}_{n-1}^1(t)=-i\left(\Delta-\frac{i}{2T_1}\right)G_{n-1}^1(t)-i\Lambda G_n^0(t). \tag{16}$$

Using these two equations and their complex-conjugates, we easily obtain, using Eq. (11), the following system of equations for elements of the density matrix:

$$\dot{W}_{10}=-i(\Delta-i\Gamma)W_{10}-\chi(W_0-W_1),$$
$$\dot{W}_{01}=i(\Delta+i\Gamma)W_{01}-\chi(W_0-W_1),$$
$$\dot{W}_1=-\chi(W_{10}+W_{01})-W_1/T_1,$$
$$\dot{W}_0=\chi(W_{10}+W_{01}), \tag{17}$$

where $\Gamma=1/2T_1$ and

$$W_{10}(t)=G_{n-1}^1(t)G_n^{*0}(t),\quad W_{01}(t)=W_{10}^*(t). \tag{18}$$

Although this system of equations resembles the system of optical Bloch equations, it has two essential distinctions. First, in the equations for nondiagonal elements we see $2T_1$ instead of the dephasing time T_2. This difference is due to neglect of the electron–phonon interaction. Secondly, Eqs. (17) and (12) yield not the conservation law $\dot{W}_0+\dot{W}_1=0$, which follows from the Bloch equations, but the relation

$$\dot{W}_0+\dot{W}_1=-\frac{W_1}{T_1}=-\sum_{\mathbf{k}}\dot{W}_{\mathbf{k}}. \tag{19}$$

By taking into account Eq. (3), we obtain a very simple expression for the two-photon correlator measured in the start-stop regime:

$$s(t)=W_1/T_1, \tag{20}$$

where the time-dependent probability $W_1(t)$ is derived from equation system (17).

3. FULL TWO-PHOTON CORRELATOR

The two-photon correlator is measured in the start-stop regime under the condition that one photon immediately follows another. But one can also measure the so-called full correlator, i.e., ignore the latter condition and detect *all* pairs of photons separated by the time interval t_0. Let us again take as an illustration Fig. 1. Whereas in the start-stop regime we detect only the pairs (2, 3) and (12, 13), in measuring the full two-photon correlator we should additionally detect the pair (4, 6). In the interval between photons 4 and 6 the atom has emitted photon 5. The two-photon correlator taking into account all photon pairs separated by a given time interval is termed the full two-photon correlator. The count rate of such photon pairs will be denoted by $p(t)$. Let us derive the formula for the full two-photon correlator.

It is obvious that, in counting all photon pairs detected by two PMTs per unit time, we measure the probability

$$p(t)=\rho_{11}(t)/T_1, \tag{21}$$

where $\rho_{11}(t)$ is the probability of detecting the atom in the excited state. In what follows, we will find out that such a probability can be derived from an equation system similar

to the Bloch equations. We can also suggest a simple justification of Eq. (21) on the qualitative level. Let us first consider the steady case over a sufficiently long time interval, when the result is independent of the length of the interval. Then we have the time-independent probability of detecting the atom in the excited state, $\rho_{11}(\infty)$. By multiplying this probability by the total number N of atoms, we obtain the number of atoms in the excited state, $N_1(\infty)=N\rho_{11}(\infty)$. The atoms spontaneously emit photons and get to the ground state. In the steady state, laser pumping compensates for the loss of excited atoms due to generation of photons. Therefore, the flow of emitted photons is constant, and the number of photons emitted per unit time is $N_1(\infty)/T_1$. Therefore the probability of photon emission by one atom in the steady state is

$$p(\infty)=\rho_{11}(\infty)/T_1. \tag{22}$$

It is obvious that, if the population of excited atoms is a function of time, the probability of photon emission is also time-dependent, in accordance with Eq. (21). Therefore, now we are confronting the problem of deriving equations for the new probability ρ_{11} of detecting the atom in the excited state irrespective of how many times it has returned to the ground state since the time $t=0$. It is obvious that the numerator on the right of Eq. (21) should take into account the number of all photons separated by time interval t_0, but not only those which immediately follow one another, as it is done in the start-stop regime. In order to calculate this probability, we have to reconsider the infinite chain of equations for amplitudes (9), i.e., the full set of states in Eq. (8).

Consider Eq. (9) for the probability amplitudes. Now we do not cut off the equation chain and also write down the equations of higher orders. Then we have the following system:

$$\begin{aligned}
&\dot G^0=-i\Lambda^* G^1,\\
&\dot G^1=-i\Delta G^1-i\left[\Lambda G^0+\sum_{\mathbf{k}}\lambda_{\mathbf{k}}G^0_{\mathbf{k}}\right],\\
&\dot G^0_{\mathbf{k}}=-i\Delta_{\mathbf{k}}G^0_{\mathbf{k}}-i[\lambda^*_{\mathbf{k}}G^1+\Lambda^{*\prime}G^1_{\mathbf{k}}],\\
&\dot G^1_{\mathbf{k}}=-i(\Delta+\Delta_{\mathbf{k}})G^1_{\mathbf{k}}-i\left[\Lambda'_{\mathbf{k}}G^0_{\mathbf{k}}+\sum_{\mathbf{k}'}\lambda_{\mathbf{k}'}G^0_{\mathbf{k}\mathbf{k}'}+\lambda_{\mathbf{k}}\sqrt{2}\underline{G^0_{2\mathbf{k}}}\right],
\end{aligned} \tag{23}$$

$$\begin{aligned}
&\underline{\dot G^0_{2\mathbf{k}}}=\ldots,\\
&\dot G^0_{\mathbf{k}\mathbf{k}'}=-i(\Delta_{\mathbf{k}}+\Delta_{\mathbf{k}'})G^0_{\mathbf{k}\mathbf{k}'}-i[\lambda^*_{\mathbf{k}}G^1_{\mathbf{k}'}+\lambda^*_{\mathbf{k}'}G^1_{\mathbf{k}}+\Lambda^{*\prime\prime}G^1_{\mathbf{k}\mathbf{k}'}],\\
&\dot G^1_{\mathbf{k}\mathbf{k}'}=-i(\Delta+\Delta_{\mathbf{k}}+\Delta_{\mathbf{k}'})G^1_{\mathbf{k}\mathbf{k}'}-i\Bigg[\lambda''G^0_{\mathbf{k}\mathbf{k}'}\\
&\qquad+\sum_{\mathbf{k}'}\lambda_{\mathbf{k}'}G^0_{\mathbf{k}\mathbf{k}'\mathbf{k}''}+\lambda_{\mathbf{k}}\sqrt{2}\underline{G^0_{2\mathbf{k}\mathbf{k}'}}+\lambda_{\mathbf{k}'}\sqrt{2}\underline{G^0_{\mathbf{k}2\mathbf{k}'}}\Bigg],
\end{aligned}$$

Here we use a simplified notation and omit the subscripts n, $n-1$, $n-2$, etc., of the laser mode amplitudes. All the other notation is the same as in Eq. (9). This infinite chain of equations takes into account all states in the chain (8). We use three approximations in our calculations.

The first approximation is that we neglect states in which wave vectors $\mathbf{k}, \mathbf{k}', \mathbf{k}'', \ldots$ of the spontaneously emitted photons are equal, i.e., the states with $\mathbf{k}=\mathbf{k}'$, etc. Since the variable $\mathbf{k}$ is continuous, the number of neglected states is vanishingly small in comparison with the number of remaining states. This approximation allows us to omit from the system (23) all underlined and similar terms not written explicitly.

The second approximation concerns matrix elements $\Lambda=\sqrt{n}\lambda_0$, $\Lambda'=\sqrt{n-1}\lambda_0$, $\Lambda''=\sqrt{n-2}\lambda_0, \ldots,$ depending on the number n of laser photons. Since this number is large, we assume that $\Lambda=\Lambda'=\Lambda''=\ldots$

The third approximation can be formulated as follows:

$$\begin{aligned}
&\sum_{\mathbf{k}}\lambda_{\mathbf{k}}G^0_{\mathbf{k}}=-\frac{i}{2T_1}G^1,\quad \sum_{\mathbf{k}'}\lambda_{\mathbf{k}'}G^0_{\mathbf{k}\mathbf{k}'}=-\frac{i}{2T_1}G^1_{\mathbf{k}},\\
&\sum_{\mathbf{k}''}\lambda_{\mathbf{k}''}G^0_{\mathbf{k}\mathbf{k}'\mathbf{k}''}=-\frac{i}{2T_1}G^1_{\mathbf{k}\mathbf{k}'},\ldots
\end{aligned} \tag{24}$$

We have used this approximation previously [Eq. (14)], when we analyzed the truncated hierarchy of equations (9). The second, third, etc., lines in Eq. (24) are justified in a similar manner. The approximation (24) is of fundamental importance in separating the infinite chain of coupled equations into an infinite number of decoupled pairs of equations. Thus, using all these three approximations, we transform the system of equations (23) to

$$\begin{aligned}
&\dot G^0=-i\Lambda^* G^1,\\
&\dot G^1=-i\Delta G^1-i\left[\Lambda G^0-\frac{i}{2T_1}G^1\right];\\
&\dot G^0_{\mathbf{k}}=-i\Delta_{\mathbf{k}}G^0_{\mathbf{k}}-i[\lambda^*_{\mathbf{k}}G^1+\Lambda^*G^1_{\mathbf{k}}],\\
&\dot G^1_{\mathbf{k}}=-i(\Delta+\Delta_{\mathbf{k}})G^1_{\mathbf{k}}-i\left[\Lambda G^0_{\mathbf{k}}-\frac{i}{2T_1}G^1_{\mathbf{k}}\right];\\
&\dot G^0_{\mathbf{k}\mathbf{k}'}=-i(\Delta_{\mathbf{k}}+\Delta_{\mathbf{k}'})G^0_{\mathbf{k}\mathbf{k}'}-i[\lambda^*_{\mathbf{k}}G^1_{\mathbf{k}'}+\lambda^*_{\mathbf{k}'}G^1_{\mathbf{k}}+\Lambda^*G^1_{\mathbf{k}\mathbf{k}'}],\\
&\dot G^1_{\mathbf{k}\mathbf{k}'}=-i(\Delta+\Delta_{\mathbf{k}}+\Delta_{\mathbf{k}'})G^1_{\mathbf{k}\mathbf{k}'}-i\left[\Lambda G^0_{\mathbf{k}\mathbf{k}'}-\frac{i}{2T_1}G^1_{\mathbf{k}\mathbf{k}'}\right];
\end{aligned} \tag{25}$$

The first pair of equations describes the evolution of amplitudes in the presence of the photon detected by the first PMT. The second pair of equations applies to the case when the laser field and the second photon with wave vector $\mathbf{k}$ emitted by the atom are present. The third pair of equations describes two spontaneously emitted photons $\mathbf{k}$ and $\mathbf{k}'$, etc. In analyzing the time dependence of the amplitudes in the start-stop regime, we ignored all equation pairs except the first. Now we will calculate the time dependence of the probabilities, taking into account all equations of the system (25).

We have derived the system of four equations (17) from the first two lines of Eq. (25) and the complex-conjugate equations at the end of the previous section. Let us apply this procedure to other equation pairs. Consider the second pair

of lines in Eq. (25) and their complex conjugates. Using the method developed above, we derive the following equations for the elements of the density operator:

$$\frac{d}{dt}\sum_{\mathbf{k}} G^1_{\mathbf{k}}G^{*0}_{\mathbf{k}} = -i\left(\Delta - \frac{i}{2T_1}\right)\sum_{\mathbf{k}} G^1_{\mathbf{k}}G^{*0}_{\mathbf{k}} - \chi\left(\sum_{\mathbf{k}} |G^0_{\mathbf{k}}|^2 - \sum_{\mathbf{k}} |G^1_{\mathbf{k}}|^2\right) + i\underline{\sum_{\mathbf{k}} \lambda_{\mathbf{k}}G^1_{\mathbf{k}}G^{*1}},$$

$$\frac{d}{dt}\sum_{\mathbf{k}} G^0_{\mathbf{k}}G^{*1}_{\mathbf{k}} = \left(\frac{d}{dt}\sum_{\mathbf{k}} G^1_{\mathbf{k}}G^{*0}_{\mathbf{k}}\right)^*,$$

$$\frac{d}{dt}\sum_{\mathbf{k}} |G^1_{\mathbf{k}}|^2 = -\chi\sum_{\mathbf{k}} (G^1_{\mathbf{k}}G^{*0}_{\mathbf{k}} + G^0_{\mathbf{k}}G^{*1}_{\mathbf{k}}) - \frac{1}{T_1}\sum_{\mathbf{k}} |G^1_{\mathbf{k}}|^2, \tag{26}$$

$$\frac{d}{dt}\sum_{\mathbf{k}} |G^0_{\mathbf{k}}|^2 = \chi\sum_{\mathbf{k}} (G^1_{\mathbf{k}}G^{*0}_{\mathbf{k}} + G^0_{\mathbf{k}}G^{*1}_{\mathbf{k}}) + \frac{1}{T_1}\sum_{\mathbf{k}} |G^1_{\mathbf{k}}|^2.$$

In deriving these equations, we have used the approximation (24). One can also prove the following relations:

$$\sum_{\mathbf{k}} \lambda_{\mathbf{k}}G^1_{\mathbf{k}} \simeq \frac{i}{2T_1}\frac{\Lambda}{\Omega}G^1, \quad \sum_{\mathbf{k}'} \lambda_{\mathbf{k}'}G^1_{\mathbf{k}\mathbf{k}'} \simeq \frac{i}{2T_1}\frac{\Lambda}{\Omega}G^1_{\mathbf{k}},$$

$$\sum_{\mathbf{k}''} \lambda_{\mathbf{k}''}G^1_{\mathbf{k}\mathbf{k}'\mathbf{k}''} \simeq \frac{i}{2T_1}\frac{\Lambda}{\Omega}G^1_{\mathbf{k}\mathbf{k}'}, \ldots \tag{27}$$

Allowing for these formulas and the inequality $\Lambda \ll \Omega$, we can omit the underlined term in the first line of Eq. (26). Similar omissions based on Eq. (27) can be done in equations for the density operator derived from the third and the following pairs of lines in Eq. (25). This is **the fourth approximation** which can be formulated in addition to the previous three.

Using the third pair of lines in Eq. (25) and the fourth approximation, we obtain the following system of equations:

$$\frac{d}{dt}\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} G^1_{\mathbf{k}\mathbf{k}'}G^{*0}_{\mathbf{k}\mathbf{k}'} = -i\left(\Delta - \frac{i}{2T_1}\right)\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} G^1_{\mathbf{k}\mathbf{k}'}G^{*0}_{\mathbf{k}\mathbf{k}'} - \chi\left(\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} |G^0_{\mathbf{k}\mathbf{k}'}|^2 - \frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} |G^1_{\mathbf{k}\mathbf{k}'}|^2\right),$$

$$\frac{d}{dt}\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} G^0_{\mathbf{k}\mathbf{k}'}G^{*1}_{\mathbf{k}\mathbf{k}'} = \left(\frac{d}{dt}\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} G^1_{\mathbf{k}\mathbf{k}'}G^{*0}_{\mathbf{k}\mathbf{k}'}\right)^*, \tag{28}$$

$$\frac{d}{dt}\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} |G^1_{\mathbf{k}\mathbf{k}'}|^2 = -\chi\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} (G^1_{\mathbf{k}\mathbf{k}'}G^{*0}_{\mathbf{k}\mathbf{k}'} + G^0_{\mathbf{k}\mathbf{k}'}G^{*1}_{\mathbf{k}\mathbf{k}'}) - \frac{1}{T_1}\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} |G^1_{\mathbf{k}\mathbf{k}'}|^2,$$

$$\frac{d}{dt}\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} |G^0_{\mathbf{k}\mathbf{k}'}|^2 = \chi\frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} (G^1_{\mathbf{k}\mathbf{k}'}G^{*0}_{\mathbf{k}\mathbf{k}'} + G^0_{\mathbf{k}\mathbf{k}'}G^{*1}_{\mathbf{k}\mathbf{k}'}) + \frac{1}{T_1}\sum_{\mathbf{k}\mathbf{k}'} |G^1_{\mathbf{k}}|^2.$$

On the base of the next pairs of lines in Eq. (25), one can construct groups of four equations similar to Eq. (28). The prescriptions for their derivation are straightforward.

Let us introduce the following infinite sums:

$$\rho_{00} = W_0 + \sum_{\mathbf{k}} |G^0_{\mathbf{k}}|^2 + \frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} |G^0_{\mathbf{k}\mathbf{k}'}|^2 + \ldots,$$

$$\rho_{11} = W_1 + \sum_{\mathbf{k}} |G^1_{\mathbf{k}}|^2 + \frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} |G^1_{\mathbf{k}\mathbf{k}'}|^2 + \ldots,$$

$$\rho_{10} = W_{10} + \sum_{\mathbf{k}} G^1_{\mathbf{k}}G^{*0}_{\mathbf{k}} + \frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} G^1_{\mathbf{k}\mathbf{k}'}G^{*0}_{\mathbf{k}\mathbf{k}'} + \ldots, \tag{29}$$

$$\rho_{01} = W_{01} + \sum_{\mathbf{k}} G^0_{\mathbf{k}}G^{*1}_{\mathbf{k}} + \frac{1}{2!}\sum_{\mathbf{k}\mathbf{k}'} G^0_{\mathbf{k}\mathbf{k}'}G^{*1}_{\mathbf{k}\mathbf{k}'} + \ldots$$

It is obvious that each of the four new elements ρ_{00}, ρ_{11}, ρ_{10}, and ρ_{01} is the trace of the full density matrix of the ''atom + field'' system with respect to the quantum numbers of the spontaneously emitted photons. The new matrix elements are elements of the atomic density matrix, which no longer depend on the indices of field modes. Using this density operator, we can calculate the mean value of any operator acting on the dynamic variables of the atom.

Now let us derive equations which are satisfied by the infinite sums (29). To this end, let us add the first, second, third, and fourth lines in Eqs. (17), (26), (28), etc., respectively. As a result, we obtain the following four equations for the elements of the atomic density matrix:

$$\dot{\rho}_{10} = -i(\Delta - i\Gamma)\rho_{10} - \chi(\rho_0 - \rho_1),$$

$$\dot{\rho}_{01} = i(\Delta + i\Gamma)\rho_{01} - \chi(\rho_0 - \rho_1), \tag{30}$$

$$\dot{\rho}_{11} = -\chi(\rho_{10} + \rho_{01}) - \rho_{11}/T_1,$$

$$\dot{\rho}_{00} = \chi(\rho_{10} + \rho_{01}) + \rho_{11}/T_1,$$

where $\Gamma = 1/2T_1$. This is the system of optical Bloch equations for a two-level atom which takes into account spontaneous light emission of this atom, i.e., T_1-relaxation, but neglects the electron–phonon interaction. The last term in the last line is the full two-photon correlator. It should be calculated by solving the system of equations (30).

4. EFFECT OF PHONONS AND TUNNELING EXCITATIONS ON CORRELATOR EQUATIONS

Of course, a two-level atom or molecule in a solid lattice interacts with phonons. There are tunneling systems in polymers and glasses in addition to phonons. Anderson *et al.*[16] and Phillips[17] suggested modeling such entities by two-level systems. Hence the question arises as to how the interaction between optically active electrons of an impurity on one side, and phonons and tunneling systems on the other should affect the systems (17) and (30) for calculating two-photon correlators of either type.

In order to clarify this issue, let us consider an equation for the density operator

$$\dot{\hat{\rho}}=\frac{1}{i\hbar}[H,\hat{\rho}], \tag{31}$$

which includes the Hamiltonian of the full system incorporating the electrons of the chromophore, the phonons, the excitations of the two-level systems (tunnelons), and photons. It can be expressed as

$$H=H_{\rm ch}+H_{\perp}+\Lambda, \tag{32}$$

where $H_{\perp}$ is the Hamiltonian due to the transverse electromagnetic field and Λ is the operator of the interaction between this field and the chromophore. To begin with, we assume that we know the solutions to the Schrödinger equation with the Hamiltonian $H_{\rm ch}$:

$$H_{\rm ch}|0\rangle|a\rangle=\hbar\Omega_a|0\rangle|a\rangle, \quad H_{\rm ch}|1\rangle|b\rangle=\hbar(\Omega+\Omega_b)|1\rangle|b\rangle, \tag{33}$$

where Ω is the chromophore excitation frequency, $|0\rangle$ and $|1\rangle$ are its wave functions, and $|a\rangle$ and $|b\rangle$ are functions corresponding to the states of the phonon system and two-level systems coupled to the chromophore in the ground and excited electron states, respectively. The eigenfunctions of the system including the noninteracting chromophore and electromagnetic field are direct products of eigenfunctions of the chromophore and field:

$$|a)=|0\rangle|n\rangle|a\rangle, \quad |b)=|1\rangle|n'\rangle|b\rangle, \tag{34}$$

where n is the number of photons in the active laser field mode. If we neglect spontaneous emission of light, the operator equation (31) in the basis of the functions (34) reads as follows:

$$\begin{aligned}
&\dot{\rho}_{ba}=-i\omega_{ba}\rho_{ba}-i\sum_{a'}\Lambda_{ba'}\rho_{a'a}+i\sum_{b'}\rho_{bb'}\Lambda_{b'a},\\
&\dot{\rho}_{ab}=\dot{\rho}^{*}_{ba},\\
&\dot{\rho}_{bb'}=-i\omega_{bb'}\rho_{bb'}-i\sum_{a}(\Lambda_{ba}\rho_{ab'}-\rho_{ba}\Lambda_{ab'}),\\
&\dot{\rho}_{aa'}=-i\omega_{aa'}\rho_{aa'}-i\sum_{b}(\Lambda_{ab}\rho_{ba'}-\rho_{ab}\Lambda_{ba'}),
\end{aligned} \tag{35}$$

where the frequencies with two subscripts are differences between the corresponding frequencies of the unperturbed Hamiltonian. It is clear that including the interaction with phonons and two-level systems inevitably leads to an infinite system of equations. If the spontaneous emission of light is taken into account, the system (35) becomes even more complicated. Such a system was presented in the previous publication.[15] Nonetheless, this infinite system of equations can be successfully used in analysis of specific effects, such as the shapes of the spectral holes burnt in optical bands by laser light or the signals of the femtosecond photon echo.[18,19] It was also shown in the previous publication[15] that, if the phononless optical line dominates over a real optical range, the infinite set of equations for such a system can be reduced to four Bloch equations (30) with

$$\Gamma\equiv\frac{1}{T_2}=\frac{\gamma(T)}{2}+\frac{1}{2T_1}. \tag{36}$$

Here the temperature-dependent FWHM $\gamma(T)$, which is called the pure dephasing constant, is completely determined by the quadratic electron–phonon and electron–tunnelon interactions. Therefore, by substituting in Eqs. (17) and (30) the constant Γ given by Eq. (36), we take into account the effect of phonons and tunnelons on the phononless line. This case is of considerable importance for the spectroscopy of isolated molecules, whose spectral lines are usually phononless.

5. RELATION BETWEEN TWO TYPES OF CORRELATORS

Let us determine the relationship between the two-photon correlators measured in different fashion. They are given by the formulas $p(t)=\rho_{11}(t)/T_1$ and $s(t)=W_1(t)/T_1$, respectively, where $\rho_{11}(t)$ and $W_1(t)$ are solutions of Eqs. (30) and (17), respectively. It is easier to determine the relation between Laplace components of these probabilities. First let us consider the probability corresponding to the two-photon correlator. By taking the formulas

$$(G(t))_{\omega}=G(\omega)=\int_0^{\infty}G(t)e^{it(\omega+i0)}dt, \tag{37}$$

$$(\dot{G}(t))_{\omega}=-G(t=0)-i(\omega+i0)G(\omega) \tag{38}$$

instead of Eq. (30) for the time components, we obtain an equation system for the Laplace components of the density matrix:

$$\begin{aligned}
&[i\omega-i(\Delta-i\Gamma)]\rho_{10}=\chi(\rho_{00}-\rho_{11}),\\
&[i\omega+i(\Delta+i\Gamma)]\rho_{01}=\chi(\rho_{00}-\rho_{11}),\\
&(i\omega-1/T_1)\rho_{11}=\chi(\rho_{10}+\rho_{01}),\\
&i\omega\rho_{00}+(1/T_1)\rho_{11}=-\chi(\rho_{10}+\rho_{01})-\rho_{00}(0).
\end{aligned} \tag{39}$$

By expressing the nondiagonal elements of the density matrix in terms of the diagonal elements, we obtain the following equations for the diagonal elements:

$$\begin{aligned}
&(i\omega-1/T_1-k)\rho_{11}+k\rho_{00}=0,\\
&(1/T_1+k)\rho_{11}+(i\omega-k)\rho_{00}=-1,
\end{aligned} \tag{40}$$

where

$$k(\omega)=2i\chi^2\frac{\omega+i\Gamma}{(\omega+i\Gamma)^2-\Delta^2}. \tag{41}$$

By solving equations (40) for ρ_{11} and taking into account the formula $p(t)=\rho_{11}(t)/T_1$, we obtain

$$p(\omega)=\frac{\rho_{11}(\omega)}{T_1}=\frac{k}{T_1D_p}, \tag{42}$$

where the determinant of the system (40) is given by the formula

$$D_p=i\omega(i\omega-1/T_1-2k). \tag{43}$$

Now let us consider the probability measured in the start-stop regime. As was noted above, it is determined by

$s(t)=W_1(t)/T_1$, where $W_1(t)$ is derived from Eq. (17). By performing the Laplace transformation of this equation system, we get

$$\begin{aligned}
&[i\omega-i(\Delta-i\Gamma)]W_{10}=\chi(W_0-W_1),\\
&[i\omega+i(\Delta+i\Gamma)]W_{01}=\chi(W_0-W_1),\\
&(i\omega-1/T_1)W_1=\chi(W_{10}+W_{01}),\\
&i\omega W_0=-\chi(W_{10}+W_{01})-W_0(0).
\end{aligned} \tag{44}$$

After eliminating from this system the functions W_{10} and W_{01}, we find

$$\begin{aligned}
&(i\omega-1/T_1-k)W_1+kW_0=0,\\
&kW_1+(i\omega-k)W_0=-1.
\end{aligned} \tag{45}$$

By solving these equations we have

$$s(\omega)=\frac{k}{T_1D_s}, \tag{46}$$

where

$$D_s=D_p+k/T_1. \tag{47}$$

Using Eqs. (42), (43) and (46), (47), we easily find $p-s=ps$. From this relationship we derive the following equation for the Laplace transform of the full correlator:

$$p(\omega)=s(\omega)+s(\omega)p(\omega). \tag{48}$$

After taking into consideration the formula

$$\int_{-\infty}^{\infty}s(\omega)p(\omega)e^{-it\omega}\frac{d\omega}{2\pi}=\int_0^t s(t-x)p(x)dx \tag{49}$$

and performing the inverse Laplace transform of both sides of Eq. (48), we easily obtain the following equation relating the full two-photon correlator $p(t)$ to the correlator $s(t)$ measured in the start-stop regime:

$$p(t)=s(t)+\int_0^t s(t-x)p(x)dx. \tag{50}$$

An iterative solution of this equation yields the probability $p(t)$ in the form of a power series in the correlator s. The physical meaning of this expansion is transparent. The term linear in s corresponds to the start-stop regime, when the atom has emitted no photon in the interval between the two detected photons. The term

$$p_1(t)=\int_0^t s(t-x)s(x)dx \tag{51}$$

corresponds to the situation when the atom emits one undetected photon in the interval between two detected photons, i.e., photon 5 between photons 4 and 6. in Fig. 1. The term cubic in s corresponds to the situation with two undetected photons in the interval t, etc. Equations (50) and (51) determine the relationship between the two kinds of two-photon correlators.

6. TWO-PHOTON CORRELATORS AS FUNCTIONS OF TIME AND LASER LIGHT FREQUENCY

Let us consider first the full two-photon correlator as a function of time. By substituting expression (43) in Eq. (42), we obtain the following expression for the Laplace transform of the desired probability:

$$p(\omega)=\frac{2i\chi^2}{T_1}\frac{\omega+i\Gamma}{Q_p(\omega)}, \tag{52}$$

where

$$\begin{aligned}
Q_p(\omega)&=\omega\{4\chi^2(\omega+i\Gamma)-(\omega+i/T_1)[(\omega+i\Gamma)^2-\Delta^2]\}\\
&=(\omega-\omega_1)(\omega-\omega_2)(\omega-\omega_3)(\omega-\omega_4).
\end{aligned} \tag{53}$$

Here ω_j are the roots of the equation $Q_p(\omega)=0$. These four roots are the poles of the desired function, so it can be expressed as a sum of pole terms:

$$p(\omega)=\sum_{j=1}^{4}\frac{p_j}{\omega-\omega_j}, \tag{54}$$

where

$$p_j=[(\omega-\omega_j)p(\omega)]_{\omega=\omega_j}. \tag{55}$$

The inverse Laplace transform of Eq. (54) yields the following expression for the full two-photon correlator:

$$p(t)=-i\sum_{j=1}^{4}p_j\exp(-i\omega_j t). \tag{56}$$

One conclusion about the time dependence of this correlator can be drawn using the general formula (56). Since one root of the polynomial $Q_p(\omega)$ is zero, one term in the sum in Eq. (56) is time-independent, so the full two-photon correlator, unlike the correlator measured in the start-stop regime, does not tend to zero as the time tends to infinity.

The expression for the time-dependent two-photon correlator measured in the start-stop regime can be derived similarly to that for the full two-photon correlator given above. By substituting Eq. (41) in (46), we obtain the following expression for the Laplace transform of the probability:

$$s(\omega)=\frac{2i\chi^2}{T_1}\frac{\omega+i\Gamma}{Q_s(\omega)}, \tag{57}$$

where

$$\begin{aligned}
Q_s(\omega)&=\left(\omega+\frac{i}{2T_1}\right)4\chi^2(\omega+i\Gamma)\\
&\quad-\omega\left(\omega+\frac{i}{T_1}\right)[(\omega+i\Gamma)^2-\Delta^2]\\
&=(\omega-\nu_1)(\omega-\nu_2)(\omega-\nu_3)(\omega-\nu_4).
\end{aligned} \tag{58}$$

Here ν_j are the roots of the equation $Q_s(\omega)=0$. The two-photon correlator measured in the start-stop regime is expressed in terms of these roots as follows:

$$s(t)=-i\sum_{j=1}^{4}s_j\exp(-i\nu_j t), \tag{59}$$

where

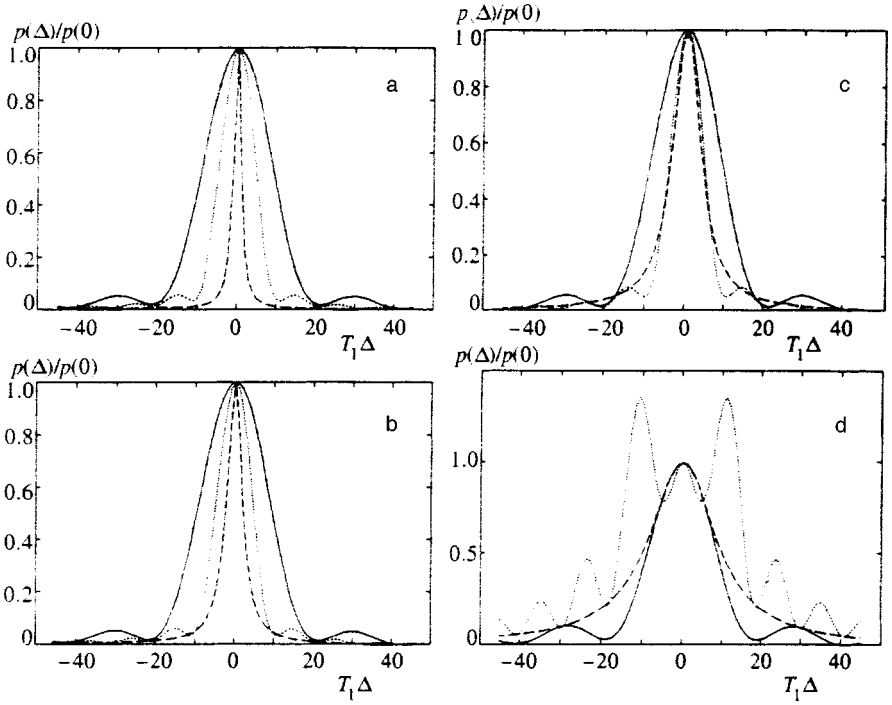

FIG. 2. Full two-photon correlator as a function of the detuning $T_1\Delta$ at various pumping intensities: (a) $\chi T_1=0.1$; (b) 1; (c) 2; (d) 5; and different times: $t/T_1=0.3$ (solid line); 0.6 (dotted line); 5 (dashed line).

$$s_j=[(\omega-\nu_j)s(\omega)]_{\omega=\nu_j}. \tag{60}$$

Equations (55), (56) and (59), (60) allow us to calculate both correlators as functions of (1) the time interval t between two detected photons, (2) the detuning Δ, and (3) the pumping intensity χ. The correlators can be calculated numerically at arbitrary values of relaxation constants $1/T_1$ and Γ. But in two specific cases, namely, at $\Delta=0$ and $\Gamma=1/T_1=\gamma$, fairly simple analytic expression convenient for analysis can be obtained.

The correlator $p(t)$ as a function of the detuning Δ yields the absorption line shape at different times t. Consider this dependence in the specific case $\Gamma=1/T_1=\gamma$. One can easily check that the equation $Q_p(\omega)=0$ can be solved analytically in this specific case and has the following four roots:

$$\omega_1=-i0, \quad \omega_2=-i\gamma, \quad \omega_3=-i\gamma+Q, \quad \omega_4=-i\gamma-Q, \tag{61}$$

where

$$Q=\sqrt{\Delta^2+4\chi^2}. \tag{62}$$

In this case, we can derive the following expression for the Laplace transform of the full two-photon correlator from Eqs. (55) and (56):

$$p(\omega)=\frac{2i\chi^2\gamma}{Q^2+\gamma^2}\left[\frac{1}{\omega+i0}-\frac{1}{2}\left(1-i\frac{\gamma}{Q}\right)\frac{1}{\omega+Q+i\gamma}-\frac{1}{2}\left(1+i\frac{\gamma}{Q}\right)\frac{1}{\omega-Q+i\gamma}\right]. \tag{63}$$

After the inverse Laplace transform, the pole of each term yields an exponent. Therefore, the full two-photon correlator is expressed as

$$p(t)=\frac{2\chi^2\gamma}{Q^2+\gamma^2}\left[1-\exp(-t\gamma)\left(\cos(Qt)+\frac{\gamma}{Q}\sin(Qt)\right)\right]. \tag{64}$$

Unlike the correlator measured in the start-stop regime, which vanishes as the time tends to infinity, the asymptotic limit of the full two-photon correlator is

$$p(\Delta)=\frac{2\chi^2\gamma}{\Delta^2+4\chi^2+\gamma^2}, \tag{65}$$

which describes the absorption line shape as a function of detuning. This line is a Lorentzian with FWHM given by

$$\Delta\omega_{1/2}=2\sqrt{4\chi^2+\gamma^2}, \tag{66}$$

which is a function of the pumping intensity χ^2.

The two-photon correlator allows one to analyze the absorption line shape at short times t. At this point, we have the line shape as a function of time. This effect is illustrated by Fig. 2. The three curves in each graph demonstrate how the absorption lines transform with time. They change from lines of complex shapes with central quasi-Gaussian peaks (at short times) to Lorentzians (at long times). The latter conclusion, however, applies only to the case of a low pumping power. By comparing Figs. 2a–2d, we can see how the increase in the pumping intensity changes the line shape. At high pumping rates, the Lorentzian line splits owing to the Stark effect caused by the electric field of the electromagnetic wave.

Now let us consider the time dependence of the full two-photon correlator and compare it with the time dependence of the two-photon correlator measured in the start-stop regime. We calculate the time dependence for the case of tun-

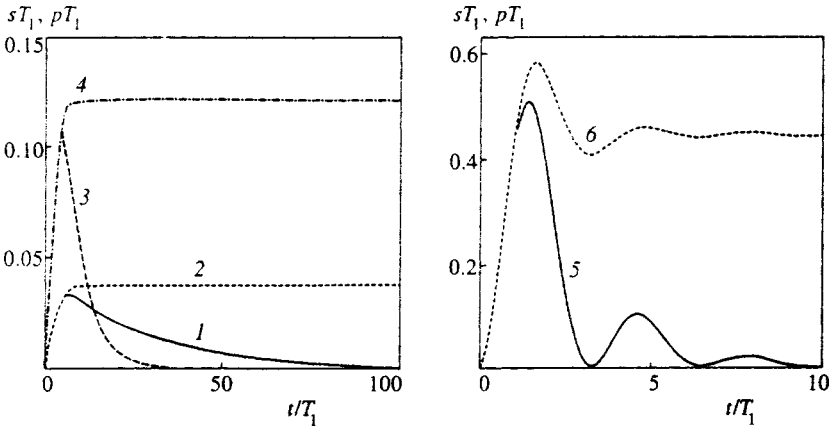

FIG. 3. Two-photon correlators $s(t)$ (curves *1*, *3*, and *5*) and $p(t)$ (curves *2*, *4*, and *6*) as functions of time between two photon detection events. $\Delta=0$; (*1*, *2*) $\chi T_1=0.1$; (*3*, *4*) 0.2; (*5*, *6*) 1.

ing to resonance, when we have $\Delta=0$ and $\Gamma=1/2T_1=\gamma/2$. In this case, the roots are given by the following formulas:

$$\nu_{1,2}=-i\frac{\gamma}{2},\quad \nu_{3,4}=-i\frac{\gamma}{2}\pm R_s,\quad R_s=\sqrt{4\chi^2-\frac{\gamma^2}{4}},$$

$$\omega_1=-i0,\quad \omega_2=-i\frac{\gamma}{2},\quad \omega_{3,4}=-i\frac{3\gamma}{2}\pm R_p, \tag{67}$$

$$R_p=\sqrt{4\chi^2-\frac{\gamma^2}{16}}.$$

The pumping intensity χ at which the functions R change from imaginary to real is different for correlators s and p. Hence the criterion for low pumping is different for the two correlators because the systems of equations (39) and (44) have different relaxation matrices.

Given the roots ν_j, we can obtain the formula for the correlator measured in the start-stop regime. By substituting these roots in Eqs. (59) and (60), we easily derive the formula for the required correlator:

$$s(t)=\frac{2\chi^2\gamma}{R_s^2}\exp\left(-\frac{\gamma}{2}t\right)[1-\cos(R_s t)]. \tag{68}$$

The roots ω_j in Eq. (67) yield the full two-photon correlator for the specific case under consideration. By substituting these roots in Eqs. (55) and (56), we obtain the formula for this correlator:

$$p(t)=\frac{2\chi^2\gamma}{R_p^2+\gamma_p^2}\left[1-\exp(-\gamma_p t)\left(\cos(R_p t)+\frac{\gamma_p}{R_p}\sin(R_p t)\right)\right], \tag{69}$$

where $\gamma_p=3\gamma/2$. This formula is very similar to Eq. (64) derived for another specific case. Figure 3 shows the curves of the start-stop correlator and full two-photon correlator versus time. The higher the pumping power, the faster the correlator rises on the initial stage. The drop in the two-photon correlator $s(t)$ at large times is also determined by the pumping intensity.

7. EFFECT OF A TRIPLET LEVEL ON THE FULL TWO-PHOTON CORRELATOR

Almost all complex organic molecules have sets of triplet levels in addition to singlet electron levels. The ground electron state of most organic molecules has spin zero, i.e., it is a singlet level, and the lowest triplet level is placed, as a rule, below the first excited singlet state, as shown in Fig. 4.

Although direct optical transitions between singlet and triplet levels are almost completely forbidden, the triplet level, nonetheless, has a considerable effect on the 0–1 transition between singlet levels because the probability of nonradiating intersystem crossing, i.e., a transition from level 1 to 2, is quite considerable. In many molecules, the constant Γ_{TS} which determines the rate of this transition is comparable to $1/T_1$ or even larger. Therefore a notable fraction of molecules transfer to the triplet level 2 without emitting a photon under cw laser radiation. The transition between the triplet level 2 and singlet level 0 is very slow. For this reason, a considerable fraction of molecules accumulate on the triplet level and thus are ''out of the play,'' i.e., they no longer absorb light. The typical rate constants for the transition shown in Fig. 4 are as follows:

$$1/T_1\simeq 10^9-10^8\ \mathrm{s}^{-1},\quad \Gamma_{TS}\simeq 10^{11}-10^{10}\ \mathrm{s}^{-1},$$
$$\gamma_{ST}\simeq 10^6-10^0\ \mathrm{s}^{-1}. \tag{70}$$

Note that the capital and small letters label relatively large and small rate constants, respectively. In the spectroscopy of single molecules, the presence of triplet levels radically changes both the time and frequency dependence of the two-photon correlator. These effects will be discussed in this section of the paper.

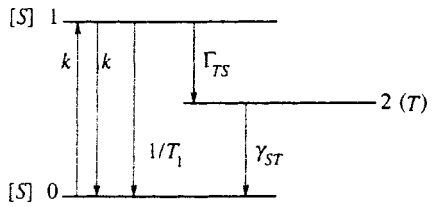

FIG. 4. Diagram of typical configuration of electron levels in a chromophore molecule and the transition rate constants.

Since the triplet level is optically inactive, the system of optical Bloch equations derived previously can be generalized to take into account these levels:

$$\begin{aligned}
&\dot{\rho}_{10}=-i(\Delta-i/T_2)\rho_{10}-\chi(\rho_0-\rho_1),\\
&\dot{\rho}_{01}=\dot{\rho}_{10}^{*},\\
&\dot{\rho}_1=-(1/T_1+\Gamma_{TS})\rho_1-\chi(\rho_{10}+\rho_{01}),\\
&\dot{\rho}_0=\rho_1/T_1+\chi(\rho_{10}+\rho_{01})+\gamma_{ST}\rho_2,\\
&\dot{\rho}_2=\Gamma_{TS}\rho_1-\gamma_{ST}\rho_2.
\end{aligned} \tag{71}$$

Here we have introduced the traditional notation for the dephasing rate by replacing Γ with $1/T_2$. The fifth line describes the population of the triplet level, and the third and fourth lines take into account the drain to the triplet level and molecules returning from this level.

The determinant of this equation system has five roots, so it can be solved only by numerical methods. However, when the pumping intensity is low in comparison with the dephasing rate $1/T_2$, we can set $\dot{\rho}_{01}=\dot{\rho}_{10}=0$ in the equations for the nondiagonal elements. In this case, the Bloch equations transform to kinetic equations, whose solution is notably easier, and the result is close to the exact solution. Thus, using this approximation, we replace Eq. (71) by the following system of kinetic equations:

$$\begin{aligned}
&\dot{\rho}_1=-(\Gamma+k)\rho_1+k\rho_0,\\
&\dot{\rho}_0=(1/T_1+k)\rho_1-k\rho_0+\gamma_{ST}\rho_2,\\
&\dot{\rho}_2=\Gamma_{TS}\rho_1-\gamma_{ST}\rho_2,
\end{aligned} \tag{72}$$

where

$$k=2\chi^2\frac{1/T_2}{\Delta^2+(1/T_2)^2},\quad \Gamma=\frac{1}{T_1}+\Gamma_{TS}. \tag{73}$$

This system of three kinetic equations can be solved analytically. We first perform the Laplace transforms of the left- and right-hand sides. Using the formula $(\dot{\rho}(t))_\omega=-\rho(t=0)-i(\omega+i0)\rho(\omega)$, we obtain the following algebraic equations for the Laplace components:

$$\begin{aligned}
&(i\omega-\Gamma-k)\rho_1+k\rho_0=0,\\
&(1/T_1+k)\rho_1+(i\omega-k)\rho_0+\gamma_{ST}\rho_2=-1,\\
&\Gamma_{TS}\rho_1+(i\omega-\gamma_{ST})\rho_2=0.
\end{aligned} \tag{74}$$

The determinant of this system can be expressed in terms of its roots:

$$\mathrm{Det}=-i(\omega+i0)(\omega-\omega_1)(\omega-\omega_2), \tag{75}$$

where

$$\begin{aligned}
&\omega_{1,2}=-i(\gamma_0\mp R),\quad \gamma_0=\frac{\Gamma+2k+\gamma_{ST}}{2},\\
&R=\sqrt{\left(\frac{\Gamma+2k-\gamma_{ST}}{2}\right)^2-\Gamma_{TS}k}.
\end{aligned} \tag{76}$$

By solving equations (74), we derive the following expression for the Laplace component of the full two-photon correlator:

$$\begin{aligned}
p(\omega)=\frac{\rho_1(\omega)}{T_1}&=\frac{ik}{T_1}\frac{i\omega-\gamma_{ST}}{(\omega+i0)(\omega-\omega_1)(\omega-\omega_2)}\\
&=\frac{ik}{T_1}\left[-\frac{\gamma_{ST}}{\omega_1\omega_2}\frac{1}{\omega+i0}+\frac{i\omega_1-\gamma_{ST}}{\omega_1(\omega_1-\omega_2)}\frac{1}{\omega-\omega_1}\right.\\
&\left.-\frac{i\omega_2-\gamma_{ST}}{\omega_2(\omega_1-\omega_2)}\frac{1}{\omega-\omega_2}\right].
\end{aligned} \tag{77}$$

The transition from the Laplace representation to the function of time is performed by replacing each pole with an exponential function with the exponent corresponding to the pole. Given that $\omega_{1,2}=-i(\gamma_0\mp R)$, we finally obtain an expression for the full two-photon correlator:

$$\begin{aligned}
p(t)=\frac{k}{T_1}\left[\frac{\gamma_{ST}}{\gamma_0^2-R^2}+\left(1-\frac{\gamma_{ST}}{\gamma_0-R}\right)\frac{\exp[-(\gamma_0-R)t]}{2R}\right.\\
\left.-\left(1-\frac{\gamma_{ST}}{\gamma_0+R}\right)\frac{\exp[-(\gamma_0+R)t]}{2R}\right].
\end{aligned} \tag{78}$$

If $\Gamma_{TS}=0$ holds, the triplet level is not populated as a result of optical pumping. One can easily verify that the first cofactor in parentheses on the right-hand side of Eq. (78) goes to zero.

Time dependence of the correlator. If we have $\Gamma_{TS}\neq 0$, transitions to the triplet level are allowed. In this case, the time dependence of the correlator is radically different. Consider the case of

$$\Gamma\gg k\gg\gamma_{ST}. \tag{79}$$

Since $\gamma_0+R\sim\Gamma$ and $\gamma_0-R\sim k$, the time dependence is easily predictable. The correlator, which is zero at $t=0$, first increases with time, then it drops and tends to a constant. This behavior of the two-photon correlator is illustrated by Fig. 5. The correlator increases concurrently with a rapid drop in the term proportional to $\exp[-(\gamma_0+R)t]$. This rise is almost independent of the pumping intensity k. The correlator drops to zero, concurrently with the slow decrease of $\exp[-(\gamma_0-R)t]$, and this drop is slower at lower k. This is clearly seen in Fig. 5, where the abscissa is the *logarithm* of time. Therefore the shift in the trailing edge of the correlator curve toward larger times means that the rate of transition from the singlet level 1 to triplet level 2 decreases by several orders of magnitude. The constant level in the limit of infinite time drops with the pumping power. This fact is easily understandable, given that the triplet population increases with the pumping, hence the probability of detecting a molecule in a singlet state, in particular, on level 1, decreases. The curves in Fig. 5 were obtained for the case of tuning to resonance, i.e., $\Delta=0$. In the nonresonant case, the curve shapes are basically the same.

The presence of a triplet level radically changes the distribution of emitted photons over time. Let us first consider the time distribution of photons emitted by a molecule without a triplet level. One can see in Fig. 3 that the full two-

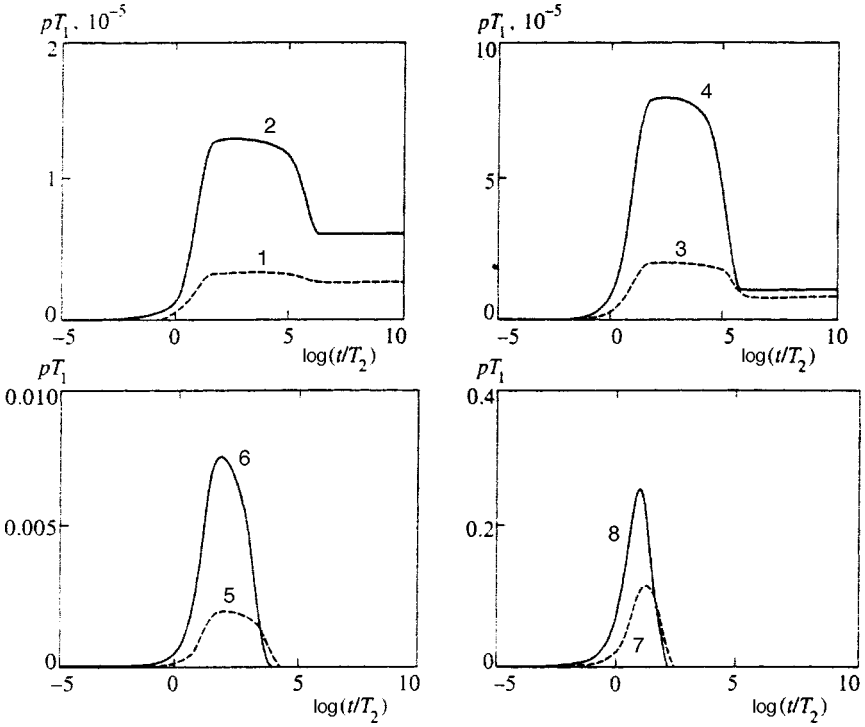

FIG. 5. Full two-photon correlator of a chromophore with a triplet level versus time. Calculations were performed using Eq. (78) at the following parameters: $T_2/T_1=10^{-2}$; $\gamma_{ST}T_2=10^{-6}$; $\Gamma_{TS}T_1=9$; (1) $\chi T_2=10^{-4}$; (2) $2\cdot10^{-4}$; (3) 10^{-3}; (4) $2\cdot10^{-3}$; (5) 10^{-2}; (6) $2\cdot10^{-2}$; (7) 10^{-1}; (8) $2\cdot10^{-1}$.

photon correlator increases with time and remains constant as the time t between the recorded photons increases. This means that two photon emission events separated by a large time are uncorrelated, i.e., they proceed independently. At small time t, on the contrary, the correlation is quite strong, namely, we can detect very few photon pairs separated by a short time interval. Therefore a sequence of photons emitted by a molecule without a triplet level is *quasi-uniform*, i.e., *without a tendency to form groups.*

The distribution of emitted photons is different if the molecule has a triplet level. Curves 5–8 in Fig. 5 indicate that light detectors should detect few pairs of photons separated by large time intervals t_1 corresponding to $p(t_1)=0$. This means that emitted photons should assemble in groups with large time intervals between them. This tendency to grouping is termed *photon bunching.* This effect is illustrated by Fig. 6, where each vertical bar denotes a propagating photon. If the average ''dimension'' of a group is t_0 and the average ''separation'' between groups is t_1 and they satisfy $t_0 \ll t_1$, the probability of detecting a photon pair with separation t_1 is very small, which is demonstrated by Fig. 5.

The effect of photon bunching can be interpreted in qualitative terms as follows. When a single chromophore molecule is exposed to cw laser light, it jumps randomly between the ground and excited singlet levels during time t_0. This process generates the photon bunch shown in Fig. 6. But if a chromophore molecule falls occasionally on the triplet level, it no longer absorbs light, and hence does not emit photons. The absence of emitted photons during the time interval t_1 is illustrated by Fig. 6. This effect of photon bunching has been observed many times in experiments with single molecules.[20]

Frequency dependence of the two-photon correlator. Equation (78) allows us to calculate the full two-photon correlator as a function of light frequency at a fixed separation t between two detected photons. First let us consider the correlator at small times. By expanding the exponential functions in Eq. (78) in powers, we obtain a simple result:

$$p(\Delta,t)\simeq \frac{t}{T_1}k=\frac{t}{T_1}\,\frac{2\chi^2/T_2}{\Delta^2+1/T_2^2}, \tag{80}$$

i.e., the correlator is proportional to the probability of absorbing a photon in unit time. This probability is described by a Lorentzian with FWHM

$$\Delta_{1/2}=2/T_2. \tag{81}$$

Now let us calculate the correlator at long times, when the correlator is constant with time. Although the correlator $p(\Delta,t=\infty)$ is small, it can be measured, as can be seen in Fig. 4. This correlator is described by the expression

$$p(\Delta,\infty)=\frac{k}{T_1}\,\frac{\gamma_{ST}}{\gamma_0^2-R^2}. \tag{82}$$

By substituting k, γ_0, and R given by Eqs. (73) and (76) in Eq. (82), we obtain

$$p(\Delta,\infty)=\frac{1}{T_1\Gamma}\,\frac{2\chi^2/T_2}{\Delta^2+(1/T_2^2)(1+2\chi^2\eta T_2\tau_T)}, \tag{83}$$

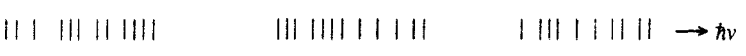

FIG. 6. Spontaneous emission of photons by a molecule with a triplet level exposed to cw laser light (photon bunching).

where

$$\eta=\frac{\gamma_{ST}+\Gamma_{TS}}{\Gamma}\approx\frac{\Gamma_{TS}}{\Gamma},\quad \tau_T=\frac{1}{\gamma_{ST}} \tag{84}$$

are the quantum yield of intersystem conversion and lifetime of the triplet level, respectively. Equation (83) describes a Lorentzian with FWHM

$$\Delta_{1/2}=\frac{2}{T_2}\sqrt{1+2\chi^2\eta T_2\tau_T}, \tag{85}$$

which is a function of the pumping intensity. By comparing Eq. (85) to the well-known formula

$$\Delta_{1/2}=\frac{2}{T_2}\sqrt{1+4\chi^2T_1T_2}$$

for a two-level chromophore, we find out that the presence of the triplet level leads to the following changes in the parameters determining the line FWHM:

$$2T_1\to\tau_T,\quad \chi^2\to\chi^2\eta. \tag{86}$$

Thus, in the presence of a triplet level the Lorentzian half width depends on the Rabi frequency when it is three orders of magnitude smaller than in the absence of a triplet level.

Summarizing our results, we can write the following three inequalities involving the pumping intensity, i.e., the Rabi frequency:

$$4\chi T_2\geqslant 1,\quad 2\chi\sqrt{T_2T_1}\geqslant 1,\quad \chi\sqrt{2\eta T_2\tau_T}\geqslant 1. \tag{87}$$

When the first condition is satisfied, which means that the pumping intensity is high, the other two conditions are satisfied automatically. In this case, the Bloch equations cannot be replaced by kinetic equations. The analysis performed in this section was based on the assumption that the inverse inequality holds:

$$4\chi T_2<1. \tag{88}$$

Since the optical dephasing life time T_2 is one or two orders of magnitude shorter than the energy relaxation time T_1, the latter condition (88) is compatible with the second and third inequalities in Eq. (87). These two conditions determine the range of pumping intensities over which the correlator Lorentzian FWHM depends on the pumping in the limit of large times. The second inequality corresponds to the case when transitions to the triplet level are forbidden for some reason, and its presence can be neglected. The third inequality defines the case when transitions from the excited singlet level to the lower triplet level are highly probable.

Equation (78) also allows one to analyze the region of intermediate times, which is a transitional region between short and long times. A numerical calculation yields the results depicted in Fig. 7. One can see that the Lorentzian is broadened at longer times, i.e., when the correlator drops and tends to a constant, as can be seen in the upper curves of Fig. 5.

8. CONCLUSIONS

The main results of the present work are Eq. (17) and (30) and the justification of their applicability to calculations of the two-photon correlator measured in the start-stop regime and the full two-photon correlator. The present numerical calculations of both correlators graphically demonstrate how various relaxation processes in single molecular impurity centers can be investigated using these correlators. The suggested approach to calculations of two-photon correlators can be generalized in order to analyze other processes, for example, slow relaxation in polymers and glasses, which leads to so-called spectral diffusion.[21,22] This problem has not been discussed in the present paper; it deserves separate study.

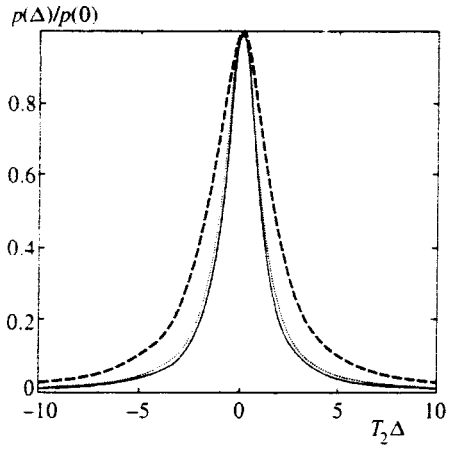

FIG. 7. Evolution of the spectral line of the full two-photon correlator: $t/T_2=10^{-5}$ (dashed line); 10^5 (solid line); 10^7 (dotted line).

The work was supported by the Russian Fund for Fundamental Research (Grant 97-02-17285) and the ''Universities of Russia'' program.

APPENDIX

If we go over to the Laplace transform in Eq. (9), we obtain the following system of equations:

$$\begin{aligned}
G_n^0&=\frac{1}{\omega+i0}+\frac{1}{\omega+i0}\Lambda^*G_{n-1}^1,\\
G_{n-1}^1&=\frac{1}{\omega-\Delta+i0}\left[\Lambda G_n^0+\sum_{\mathbf{k}}\lambda_{\mathbf{k}}G_{n-1\mathbf{k}}^0\right],\\
G_{n-1\mathbf{k}}^0&=\frac{1}{\omega-\Delta_{\mathbf{k}}+i0}[\lambda_{\mathbf{k}}^*G_{n-1}^1+\Lambda^{*\prime}G_{n-2\mathbf{k}}^1],\\
G_{n-2\mathbf{k}}^1&=\frac{1}{\omega-\Delta-\Delta_{\mathbf{k}}+i0}\left[\Lambda' G_{n-1\mathbf{k}}^0+\sum_{\mathbf{k}'}\lambda_{\mathbf{k}'}G_{n-2\mathbf{k}\mathbf{k}'}^0\right],
\end{aligned} \tag{A1}$$

Here we have used the notation

$$\Delta=\Omega-\omega_0,\quad \Delta_{\mathbf{k}}=\omega_{\mathbf{k}}-\omega_0,\quad \Lambda=\lambda_0\sqrt{n},\quad \Lambda'=\lambda_0\sqrt{n-1}, \tag{A2}$$

where

$$\lambda_0=-i\Omega\sqrt{\frac{4\pi}{\hbar\omega_0V}}\,\mathbf{e}_0\cdot\mathbf{d} \tag{A3}$$

is derived from Eq. (10) by replacing the subscript $\mathbf{k}$ of a spontaneous mode with the laser mode subscript 0.

The solution of Eq. (A.1) can be expressed in the form of a continued fraction, for example,

$$G_{n-1}^1=\frac{\Lambda G_n^0}{\omega-\Delta-\sum_{\mathbf{k}}\dfrac{|\lambda_{\mathbf{k}}|^2}{\omega-\Delta_{\mathbf{k}}-\dfrac{|\Lambda'|^2}{\omega-\Delta-\Delta_{\mathbf{k}}-\sum_{\mathbf{k}'}\ \ldots}}}. \tag{A4}$$

The matrix elements $\lambda_{\mathbf{k}}$ which do not contain the photon numbers are small, since they determine the intensity of the electromagnetic interaction, which can be treated as a perturbation. On the other hand, the matrix elements Λ and Λ' are large since they include, in accordance with Eq. (A2), the square root of the number of laser photons, which is large. The large and small matrix elements alternate in the continued fraction in Eq. (A4). It is obvious that the effect of the matrix element Λ' is compensated for by the smallness of the matrix element $\lambda_{\mathbf{k}}$. Therefore the infinite equation system (A1) can be truncated by setting $\Lambda'=0$. Then the system of equations (A1) transforms to

$$\begin{aligned} G_n^0&=\frac{1}{\omega+i0}+\frac{1}{\omega+i0}\Lambda^* G_{n-1}^1,\\ G_{n-1}^1&=\frac{1}{\omega-\Delta+i0}\left[\Lambda G_n^0+\sum_{\mathbf{k}}\lambda_{\mathbf{k}}G_{n-1\mathbf{k}}^0\right],\\ G_{n-1\mathbf{k}}^0&=\frac{1}{\omega-\Delta_{\mathbf{k}}+i0}\lambda_{\mathbf{k}}^* G_{n-1}^1. \end{aligned} \tag{A5}$$

The inverse Laplace transform yields equations (9) containing only terms written in explicitly.

*)E-mail: osad.ig@g23.relcom.ru

[1] *Spectroscopy and Excitation Dynamics of Condensed Molecular Systems*, ed. by V. Agranovich and R. Hochstrasser, North Holland, Amsterdam (1983).
[2] G. J. Small, *ibid.*, p. 515.
[3] R. I. Personov, *ibid.*, p. 555.
[4] W. E. Moerner and L. Kador, Phys. Rev. Lett. **62**, 2535 (1989).
[5] M. Orrit and J. Bernard, Phys. Rev. Lett. **65**, 2716 (1990).
[6] W. E. Moerner and Th. Basche, Adv. Mater. Opt. Electron. **32**, 457 (1993).
[7] *Progress in Optics*, vol. 35, ed. by E. Wolf, Elsevier (1996).
[8] W. P. Ambrose, Th. Basche, and W. E. Moerner, J. Chem. Phys. **95**, 7150 (1991).
[9] L. Fleury, Ph. Tamarat, B. Lounis, J. Bernard, and M. Orrit, Chem. Phys. Lett. **236**, 87 (1995).
[10] J. Kohler, J. A. J. M. Disselhorst, M. C. J. M. Donckers, E. J. J. Groenen, J. Schmidt, and W. E. Moerner, Nature (London) **363**, 242 (1993).
[11] J. Wachtrup, C. V. Borczyskowski, J. Bernard, M. Orrit, and R. Brown, Nature (London) **363**, 244 (1993).
[12] U. Wild, M. Croci, F. Guttler, M. Pirotta, and A. Renn, J. Lumin. **60–61**, 1003 (1994).
[13] Ph. Tamarat, B. Lounis, J. Bernard, M. Orrit, S. Kummer, R. Kettner, S. Mais, and Th. Basche, Phys. Rev. Lett. **75**, 1514 (1995).
[14] L. Fleury, A. Zumbusch, M. Orrit, R. Brown, and J. Bernard, J. Lumin. **56**, 15 (1993).
[15] I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **98**, 1045 (1990) [Sov. Phys. JETP **71**, 583 (1990)].
[16] P. W. Anderson, B. I. Halperin, and C. M. Varma, Philos. Mag. **25**, 1 (1972).
[17] W. A. Phillips, J. Low Temp. Phys. **7**, 351 (1972).
[18] I. S. Osad'ko and S. N. Gladenkova, Chem. Phys. Lett. **198**, 531 (1992).
[19] I. S. Osad'ko and M. V. Stashek, Zh. Éksp. Teor. Fiz. **106**, 535 (1994) [JETP **79**, 293 (1994)].
[20] J. Bernard, L. Fleury, H. Talon, and M. Orrit, J. Chem. Phys. **98**, 850 (1993).
[21] I. S. Osad'ko and N. N. Zaitsev, Chem. Phys. Lett. **257**, 82 (1996).
[22] I. S. Osad'ko, Zh. Éksp. Teor. Fiz. **109**, 805 (1996) [JETP **82**, 434 (1996)].

Translation provided by the Russian Editorial office.

Power broadening of a diffusion resonance

D. A. Shapiro*) and M. G. Stepanov

Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia
(Submitted 6 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 1632–1648 (May 1998)

We derive an expression in terms of cylinder functions for the shape of a nonlinear resonance in a two-level system with a rapidly decaying level. We show that when the natural linewidth is negligible, the square of the total width is the sum of squares of the power and diffusion widths. The traditional variational approximation yields a correct value for the full width at half maximum, but distorts the line profile. We derive a formula for the absorbed power as a function of the incident wave intensity for comparable power and diffusion broadening. The formula is found to be valid for a power width that is small or large compared to the diffusion width, and in a new intermediate domain where homogeneous saturation becomes inhomogeneous.
 [S1063-7761(98)00605-2]

1. INTRODUCTION

The subject of this paper is resonant absorption (or amplification) of a strong electromagnetic wave by ions of a plasma. Knowledge of the laws of amplification is necessary in calculating ion lasers and in optimizing the characteristics of such lasers (for example, to increase the output power or tuning range of a Raman laser). Measuring the absorption (or amplification) spectrum of waves propagating through a plasma is one of the most important plasma diagnostic techniques.

From a theoretical standpoint, the simplest and most important case (from a practical standpoint) is that of resonant interaction of an electromagnetic wave and a plasma when the wave frequency ω is close to the Bohr frequency $\omega_{21}=(E_2-E_1)/\hbar$ of the transition between intrinsic states $|1\rangle$ and $|2\rangle$ of an ion. Because of the Doppler effect, the wave interacts most effectively with ions that satisfy the resonance condition $\mathbf{k}\cdot\mathbf{v}=\omega-\omega_{21}$, where $\mathbf{k}$ is the wave vector and $\mathbf{v}$ is the ion velocity. The power absorbed from a weak electromagnetic field in the linear approximation (when changes in the ion state initiated by the wave are ignored) is proportional to the number of particles that interact with the wave. As the frequency varies, the wave resonantly interacts with ions that have a different velocity. The width of the absorption spectrum is determined by the characteristic spread of the velocities v_T of the ions, i.e., by the width of the Maxwellian velocity distribution, and is equal to kv_T. By measuring the linear absorption spectrum we can find the ion temperature.

Much more information about the plasma parameters and the relaxation processes inside an individual ion can be gathered if one employs the methods of nonlinear spectroscopy. An electromagnetic wave interacting with the plasma tends to equalize the populations of the intrinsic states of the ions. Because of the resonant nature of this interaction, the velocity distribution of the populations acquires narrow nonequilibrium structures known as Bennett dips or peaks.[1] The shape of Bennett dips can be found by the probe field technique by measuring the absorption of another electromagnetic wave as a function of frequency. The absorption spectrum of the probe wave that resonantly interacts with the same transition has a nonlinear resonance, a dip in the vicinity of the frequency at which the probe wave resonantly interacts with those ions whose populations have been partially equalized by the first wave. A similar dip emerges in the absorption spectrum of a standing wave consisting of two counterpropagating traveling waves.

The velocity of an ion in a plasma changes with time, since the ion is in the field of the other charged particles comprising the plasma. In contrast to gases, where collisions with other atoms are infrequent but in each such collision the atom dramatically changes velocity, an ion is constantly in a rapidly varying field, which leads to diffusion variations in its velocity.[2] The description of collisions as diffusion in velocity space is also possible for heavy neutral particles in a buffer gas of light particles.

The first theoretical study of the diffusion shape of the Bennett dip in a weak field as applied to ions in a plasma was Ref. 3. It was shown that because of diffusion in velocity with a coefficient D, the dip in level j acquires a width $\sqrt{D/\Gamma_j}$ equal to the characteristic variation of the ion velocity during the lifetime Γ_j^{-1} in level j. If the diffusion width exceeds the natural width, the dip in the distribution over the projection of velocity on the wave vector has a characteristic cusp at the center described by the exponential function $\exp(-|x|)$, where x is the dimensionless deviation of the velocity projection from the center of the dip. Quenching of levels by electrons and other processes described by the relaxation-constant model may add to the natural width of the dip.

In the limit of a weak electromagnetic field, the depth of Bennett dips is proportional to the wave's intensity. But the increase in the depth with wave intensity slows down as the level populations become equalized. In a strong field, the populations at the center of the resonance are essentially

equal, and the range of velocities in which the population difference decreases substantially broadens as the field intensity increases.[4] The problem of adding diffusion broadening to power broadening was studied in Ref. 5 by a variational method, and the Lorentzian contour was taken as the trial function for the dip profile. When the diffusion and power widths are comparable, the shape of the dip is some sort of average between a Lorentzian and an exponential. In Ref. 6 the shape of the dip was found in the limit of a small natural width. In the present paper we discuss this problem in greater detail.

The saturation curve, i.e., the dependence of the power absorbed from the field on the intensity of the incident wave, also makes it possible to measure a number of plasma parameters and relaxation constants of the transition. A knowledge of the analogous dependence of the gain is required to calculate the output power of a laser. Diffusion leads to a situation in which some of the ions depart from resonance as their velocities change, while other ions that have not yet absorbed (or emitted) a photon begin to participate in the resonance process. Because of this, the power absorbed from the field (or imparted to it) increases. This dependence in relatively weak fields, where the power width of the Bennett dip is much less than the diffusion width, was derived in Ref. 7 and experimentally corroborated by Apolonsky *et al.*[8] It was shown that in the presence of diffusion the absorption saturation becomes homogeneous and sets in at lower field intensities than in the absence of diffusion. The intensity at which saturations sets in proved to be proportional to the square root of the diffusion coefficient D. The case in which the diffusion and power widths of Bennett dips are comparable remained unexplored.

In Sec. 2 we give the initial equations for the ion density matrix, a classification of the processes of interaction of an ion and the field of electromagnetic waves, and the characteristic values of the main parameters. In Sec. 3 we derive an expression for the shape of the Bennett dip when there is both power broadening and diffusion broadening. In Sec. 4 we calculate the saturation curve, i.e., the power absorbed from a strong field. For this curve we give simple interpolation formulas, which make it possible to forego the assumption that the radiative width must be small. Section 5 contains a qualitative discussion of the results.

2. BASIC EQUATIONS

We normalize the Maxwellian velocity distribution of the ions to unity:

$$W(\mathbf{v})=\frac{1}{(\sqrt{\pi}v_T)^3}\exp\left(-\frac{\mathbf{v}^2}{v_T^2}\right),$$

$$v_T=\sqrt{\frac{2T}{m}},\quad \int d\mathbf{v}W(\mathbf{v})=1,$$

where T is the temperature of the ions in the plasma, and m is the ion mass. We take two excited ion states $|1\rangle$ and $|2\rangle$ with energies E_1 and E_2 ($E_1<E_2$) and examine the resonant interaction of the given two-level system with a traveling electromagnetic wave

$$\mathbf{E}(t,\mathbf{r})=\tfrac{1}{2}(\mathbf{E}_0e^{-i(\omega t-\mathbf{k}\cdot\mathbf{r})}+\text{c.c.})$$

that is offset from resonance by $\Omega=\omega-\omega_{21}\ll\omega$, where $\omega_{21}=(E_2-E_1)/\hbar$. Assuming a dipole ion–wave interaction, in the resonance approximation we have the following quantum transport equations for the density matrix in the Wigner representation:[9]

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla+\Gamma_1\right)\rho_1=\nu\hat{\mathscr{D}}\rho_1-2\,\mathrm{Re}(iG^*e^{i(\Omega t-\mathbf{k}\cdot\mathbf{r})}\rho_{21})+q_1W(\mathbf{v})+A_{21}\rho_2,\tag{1}$$

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla+\Gamma_2\right)\rho_2=\nu\hat{\mathscr{D}}\rho_2+2\,\mathrm{Re}(iG^*e^{i(\Omega t-\mathbf{k}\cdot\mathbf{r})}\rho_{21})+q_2W(\mathbf{v}),$$

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla+\Gamma_{12}\right)\rho_{21}=\nu\hat{\mathscr{D}}\rho_{21}+iGe^{-i(\Omega t-\mathbf{k}\cdot\mathbf{r})}(\rho_2-\rho_1),$$

where

$$\hat{\mathscr{D}}=\frac{\partial}{\partial v_\alpha}\Phi_{\alpha\beta}(\mathbf{v})\left(\frac{v_T^2}{2}\frac{\partial}{\partial v_\beta}+v_\beta\right);$$

$\rho_j=\langle j|\hat{\rho}|j\rangle$ and $\rho_{21}=\langle 2|\hat{\rho}|1\rangle$ are the matrix elements of the density matrix; Γ_1, Γ_2, and Γ_{12} are relaxation constants of the states $|1\rangle$ and $|2\rangle$ and of the ''coherence'' ρ_{21} of these states; $G=\mathbf{E}_0\langle 2|\hat{\mathbf{d}}|1\rangle/2\hbar$, with $\hat{\mathbf{d}}$ the dipole moment operator; $q_jW(\mathbf{v})$ is the excitation function of $|j\rangle$; A_{21} is the Einstein coefficient; the diffusion operator $\hat{\mathscr{D}}$ describes diffusion in velocity space; $\Phi_{\alpha\beta}(\mathbf{v})$ is the diffusion tensor; and ν is the transport collision rate, or the reciprocal of the time it takes a particle to change velocity (due to diffusion) by a quantity of order v_T.

Effects related to the dependence of $\Phi_{\alpha\beta}$ on $\mathbf{v}$ were discussed in Ref. 10. There it was shown that for $\nu\ll\Gamma_j$, i.e., when the diffusion width of Bennett dips along $\mathbf{k}$ is much less than v_T, we can adopt a constant and isotropic diffusion coefficient $\Phi_{\alpha\beta}(\mathbf{v})=\delta_{\alpha\beta}$, introduced into nonlinear spectroscopy by Rautian.[11] If we align the z axis with $\mathbf{k}$, then only the diffusion tensor component $\Phi_{zz}(\mathbf{v})$ averaged over the transverse velocity and taken at the longitudinal velocity corresponding to the center of the dip affects the shape of the Bennett dip. The adopted approximation makes it possible to integrate the system of equations (1) with respect to velocities transverse to $\mathbf{k}$. For the sake of brevity, we denote the longitudinal velocity projection $v_\parallel=\mathbf{k}\cdot\mathbf{v}/k$ by v.

Bearing in mind the various applications to nonlinear spectroscopy, we are interested in the steady-state solution of the system of equations (1):

$$\rho_j=r_j,\quad \rho_{21}=r_{21}e^{-i(\Omega t-\mathbf{k}\cdot\mathbf{r})},$$

$$\Gamma_1r_1=\nu\hat{\mathscr{D}}_1r_1-2\,\mathrm{Re}(iG^*r_{21})+q_1W_1(v)+A_{21}r_2,$$

$$\Gamma_2r_2=\nu\hat{\mathscr{D}}_1r_2+2\,\mathrm{Re}(iG^*r_{21})+q_2W_1(v),\tag{2}$$

$$[\Gamma_{12}-i(\Omega-kv)]r_{21}=\nu\hat{\mathscr{D}}_1r_{21}+iG(r_2-r_1),$$

$$W_1(v)=\frac{1}{\sqrt{\pi}v_T}\exp\left(-\frac{v^2}{v_T^2}\right), \quad \hat{\mathscr{T}}_1=\frac{d}{dv}\left(\frac{v_T^2}{2}\frac{d}{dv}+v\right).$$

The power absorbed from the wave per unit volume is the product of the concentration of ions absorbing a photon per unit time and the energy of the photon:

$$P(\Omega)=2\hbar\omega\int dv\ \mathrm{Re}(iG^*r_{21}). \tag{3}$$

We now present numerical values of the characteristic parameters, which can be divided into plasma parameters and the parameters of a singe ion. The ion temperature T_i characteristic of discharge is roughly 1 eV, and the concentration of ions in the plasma of an argon laser, N_i, is roughly 10^{14} cm^{-3}. The wavelength of the radiation that is in resonance with the laser transitions of a singly charged ion, λ, is approximately 5×10^{-5} cm. The relaxation constants Γ_j and Γ_{ij} are determined by the rates of relaxation processes and quenching processes in the plasma. Depending on the levels selected, these constants can vary from 10^7 to 10^9 s^{-1}. For Ar II we have $v_T\simeq2\times10^5$ cm/s, $kv_T\simeq2.5\times10^{10}$ s^{-1}, and the transport collision rate

$$\nu=\frac{16\sqrt{\pi}e^4N_i}{3m^2v_T^3}L\simeq10^7\ \mathrm{s}^{-1},$$

where e is the electron charge and L is the Coulomb logarithm. The parameters of the gas-discharge plasma have been thoroughly discussed in Ref. 9. The above values of the parameters were used in the numerical calculations. Note that all of the relaxation constants (Γ_j and Γ_{ij}) and the transport collision rate ν are small compared to the Doppler linewidth kv_T.

3. BENNETT DIP IN COMBINED POWER AND DIFFUSION BROADENING

In the absence of diffusion, the shape of the Bennett dip Δr_j is Lorentzian, and the width of the dip is the sum of the homogeneous width Γ_{12} and the power width w_F (Ref. 4):

$$\Delta r_j\propto[\Gamma_{12}^2+w_F^2+(\Omega-kv)^2]^{-1},$$

$$w_F=\sqrt{\frac{2\Gamma_{12}|G|^2}{k^2}\left(\frac{1}{\Gamma_1}+\frac{1}{\Gamma_2}-\frac{A_{21}}{\Gamma_1\Gamma_2}\right)}.$$

If $r_j(v)$ has width w, then in order to estimate the order of magnitude of the individual terms in the equation for r_j we must replace the derivative d/dv by w^{-1}. The collisional term consists of two terms, $(\nu v_T^2/2)d^2r_j/dv^2$ and $\nu(d(vr_j)/dv)$. The first describes velocity diffusion, while the second describes dynamic friction on an ion moving with respect to the entire plasma. The diffusion width w_{jD} of the Bennett dip at level j comes into play because of the equality of Γ_jr_j and the diffusion term, $(\nu v_T^2/2)d^2r_j/dv^2\sim(\nu v_T^2/2w_{jD}^2)r_j$, with the result that $w_{jD}=v_T\sqrt{\nu/2\Gamma_j}$. For moderate concentrations of charge carriers in the plasma, $\nu\ll\Gamma_j$ and $w_{jD}\ll v_T$, and we can neglect terms corresponding to dynamic friction. Due to friction, the variation of the ion velocity during the lifetime Γ_j^{-1} is small in comparison to the diffusion variation of the velocity, w_{jD}, and is of order $v_T\nu/\Gamma_j$.

Below we examine the shape of the Bennett dip in two limits:

1. The upper level is long-lived, and the lower level rapidly decaying ($\Gamma_2\ll\Gamma_1$). This situation leads, for equal rates of excitation of levels, to steady-state population inversion, and is typical of laser transitions of cw ion lasers.[9]

2. The lower level is long-lived, or $\Gamma_1\ll\Gamma_2$. A metastable lower state is used in the absorbing transition of Raman lasers.[8]

In both cases we assume that the diffusion width of the Bennett dip in the short-lived level is much less than the homogeneous or power width. Neglecting diffusion operators in the population equations of this level and for the coherence r_{21} in the system of equations (2), we obtain an ordinary second-order differential equation for the population r_2 of the long-lived level in case 1:

$$\left[1+J\left(1-\frac{A_{21}}{\Gamma_1}\right)\right]r_2=w_{2D}^2\frac{d^2r_2}{dv^2}+r_2^{(0)}+J\left(r_1^{(0)}-\frac{A_{21}}{\Gamma_1}r_2^{(1)}\right), \tag{4}$$

where

$$J=\frac{2\Gamma_{12}|G|^2}{\Gamma_2[\Gamma_{12}^2+2\Gamma_{12}|G|^2/\Gamma_1+(\Omega-kv)^2]},$$

and $r_j^{(0)}$ is the population of level j in the absence of a field. The shape of the Bennett dip is described by the function $y=(r_2-r_2^{(0)})/(r_1^{(0)}-r_2^{(0)})$, which obeys the equation

$$\left(1+\frac{A^2}{W^2+x^2}\right)y=\frac{d^2y}{dx^2}+\frac{A^2}{W^2+x^2}, \tag{5}$$

where

$$x=\frac{\Omega-kv}{kw_{2D}}, \quad A=\frac{w_F}{w_{2D}}, \quad W=\frac{(\Gamma_{12}^2+2\Gamma_{12}|G|^2/\Gamma_1)^{1/2}}{kw_{2D}}.$$

Note that although $|G|^2$ is present in the expression for W, the latter is not simply the ratio A of the power width to the diffusion width. For instance, in a strong field, $W\simeq(\Gamma_2/\Gamma_1)^{1/2}A\ll A$. The exact solutions of the homogeneous equation can be expressed in terms of spheroidal functions.[12] These functions, however, have been rather poorly studied and no integral representations are known for them.

In case 2, reasoning along similar lines, we have

$$\left[1+J\left(1-\frac{A_{21}}{\Gamma_2}\right)\right]r_1=w_{1D}^2\frac{d^2r_1}{dv^2}+r_1^{(0)}+J\left(1-\frac{A_{21}}{\Gamma_2}\right)r_2^{(0)}, \tag{6}$$

where the expression for J can be obtained from its counterpart in case 1 via the interchange $1\leftrightarrow2$. If we interchange the indices in the expressions for x, y, A, and W, we again arrive at Eq. (5).

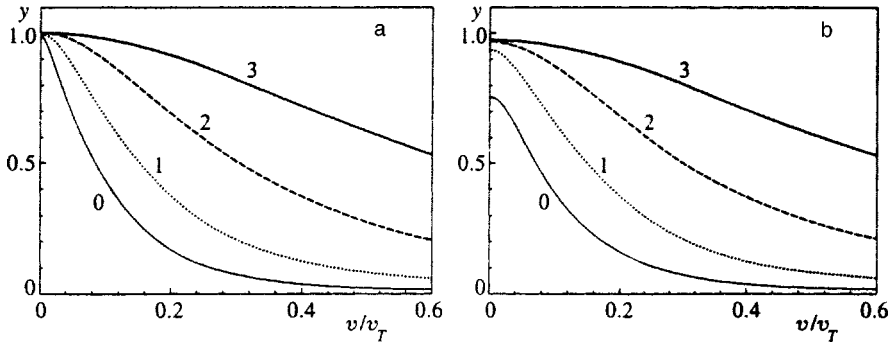

FIG. 1. The Bennett dip $y(x)$: (a) the solution (10), and (b) the results of numerical calculations. $\Gamma_<=10^{-3}kv_T$ and $\Gamma_>=4\times10^{-2}kv_T$, where $\Gamma_<$ ($\Gamma_>$) is the smaller (larger) of the two quantities Γ_1 and Γ_2. $\Gamma_{12}=(\Gamma_1+\Gamma_2)/2$, $\nu=10^{-5}kv_T$, and $\Omega=0$. The curves $n=0,1,2,3$ correspond to the different values of $|G|=10^{n/3-2}kv_T$.

Equation (5) has a simple physical meaning: there is a reservoir of particles and a set of states that differ in the value of the parameter x. In these states, a particle experiences diffusion in x and decay (the unity on the left-hand side of the equation). The external field resonantly interacts with particles, initiating transitions between the states x and the reservoir; the resonance width is W. There is a nonvanishing probability of arrival or departure of a particle in state x, and this probability is $A^2/(W^2+x^2)$.

We now analyze Eq. (5). If either A or W is large ($W\gg1$ or $A\gg1$), the width of the Bennett dip exceeds the diffusion width, and we can ignore the derivative d^2y/dx^2 in the equation, so that the Bennett dip becomes Lorentzian:

$$y=\frac{A^2}{W^2+A^2+x^2}. \tag{7}$$

In weak fields, $A\ll W\sim1$, the dip is a convolution of Lorentzian and exponential contours:

$$y=\frac{1}{2}e^{-|x|}*\frac{A^2}{W^2+x^2}\equiv\frac{A^2}{2}\int dx'\,\frac{e^{-|x-x'|}}{W^2+x'^2}. \tag{8}$$

We now examine the case in which the diffusion width of the Bennett dip in the long-lived level exceeds the homogeneous width, $W\ll1$. Replacing $1/(W^2+x^2)$ by $\pi\delta(x)/W$, we obtain

$$y=\frac{d^2y}{dx^2}+\frac{\pi A^2}{W}\delta(x)[1-y(0)],$$

whose solution is

$$y(x)=\frac{A^2}{2W/\pi+A^2}e^{-|x|}. \tag{9}$$

This solution is valid for weak fields, $A\ll1$. As the field intensity grows, the depth of the dip, $y(0)$, approaches unity, and the accuracy of the approximation based on this substitution decreases. In very strong fields, $A\sim1$, the shape of the Bennett dip differs from the exponential shape $e^{-|x|}$.

Let us construct a solution of Eq. (5) for large values of x ($x\gg W$). We will then attempt to correlate the solution's behavior with the condition of solvability for small values of $x\sim W$. When $x\gg W$, we can ignore W^2 in the denominators of (5). Then the solution $y(x)$ can be expressed in terms of cylinder functions:

$$y(x)=A^2\sqrt{i|x|}\,[CK_\alpha(|x|)-S_{-3/2,\alpha}(i|x|)], \tag{10}$$

where $\alpha=\sqrt{A^2+1/4}$, $K_\alpha(z)$ and $S_{\mu,\alpha}(z)$ are the modified Bessel function of the second kind and Lommel function,[13] and C is a constant that can be found by examining the behavior of the solution (10) at small x.

When $x\sim W$, we can ignore the unity on the left-hand side of Eq. (5), $A\gg W$. The resulting inhomogeneous equation has the trivial solution $y\equiv1$. The solution of the homogeneous equation can be expressed in terms of hypergeometric series ${}_2F_1$ (Ref. 14). However, when $A^2\gg W$, the constant C can be found from the condition that the solution (10) must be regular at $x=0$:

$$C=\frac{e^{-i\alpha\pi/2}}{2^{3/2}\pi}\Gamma\left(\frac{-1/2-\alpha}{2}\right)\Gamma\left(\frac{-1/2+\alpha}{2}\right)\times\cos\left[\frac{\pi}{2}\left(\frac{3}{2}+\alpha\right)\right].$$

The solution (10) can then be written in terms of the modified Bessel function $I_\alpha(z)$ and the generalized hypergeometric series ${}_1F_2$ (Ref. 14),

$$y(x)={}_1F_2\left(1;\frac{3/2+a}{2},\frac{3/2-a}{2};\frac{x^2}{4}\right)-\sqrt{\frac{|x|}{2}}\Gamma\left(\frac{3/2+\alpha}{2}\right)\Gamma\left(\frac{3/2-\alpha}{2}\right)I_\alpha(|x|),$$

which implies that $y(0)=1$. The solution (10) contains the Lorentzian and exponential contours as limiting cases:

$$y=\begin{cases}e^{-|x|}, & A\ll1,\\ \dfrac{1}{1+(x/A)^2}, & A\gg1.\end{cases} \tag{11}$$

In weak fields, $A^2\sim W\ll1$, the Bennett dip is described by the exponential contour (9) with $y(0)\neq1$.

Figure 1 depicts examples of Bennett dips: (a) the solution (10), and (b) the numerical solutions of four coupled diffusion equations for the elements of the density matrix in (2). The range of possible velocities in (2) is infinite, so for numerical calculations the limits of this interval were taken at $v=\pm4v_T$, and the values of the density matrix at these limits were chosen according to the asymptotic behavior of the analytic Lorentzian. Figure 1 clearly shows the transition from an exponential contour to a Lorentzian contour (from curve $n=0$ to curve $n=3$). The curves $n=1,2,3$ in Fig. 1a almost coincide with the curves $n=1,2,3$ in Fig. 1b; the dif-

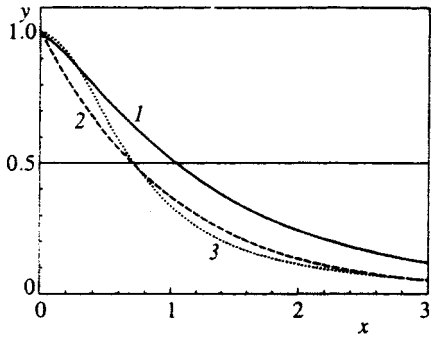

FIG. 2. The shape of the dip $y(x)$ given by formula (10) at $A=\sqrt{1/2}$ (curve *1*), the diffusion contour $\exp(-|x|)$ (curve *2*), and the Lorentzian $1/(1+(x/A)^2)$ (curve *3*).

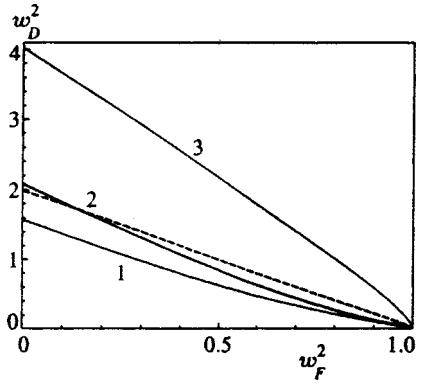

FIG. 3. Level curves for $v_{n/4}$, $n=1,2,3$, corresponding to $v_{n/4}=\sqrt{4/n-1}$. The dashed curve depicts the approximation (12) for the curve $n=2$.

ference between Figs. 1a and 1b at the center of the Bennett dip in a weak field (curve $n=0$) can be explained by the fact that only in the numerical calculations have we allowed for the finite natural width $W\neq 0$.

The type of solution $y(x)$ given by (10) depends only on the parameter $A=w_F/w_D$, where w_D is the diffusion width of the Bennett dip in the long-lived level ($w_D=w_{2D}$ in case 1 and $w_D=w_{1D}$ in case 2). Thus, the halfwidth v_L of the dip over a fraction L $(0<L<1)$ of the total height of the dip ($v_L=w_D x_L$, where $y(x_L)=L$) is a homogeneous function of w_D and w_F of order unity. This means that if w_D and w_F are increased p-fold ($p>0$), then so is v_L. In the (w_F^2,w_D^2) plane, the various level curves of v_L with the same value of L but different values of v_L, can be mapped into one another via a scale transformation centered at $(0,0)$. For instance, the halfwidth at half maximum is

$$v_{1/2}=\begin{cases} w_F\left[1+\left(\dfrac{w_D}{2w_F}\right)^2\right], & w_D\ll w_F, \\ w_D\ \ln 2, & w_D\gg w_F, \end{cases}$$

and can be approximated by the interpolation formula

$$v_{1/2}\simeq\sqrt{w_F+w_D^2/2}=w_D\sqrt{A^2+1/2}. \tag{12}$$

Note that the simplicity of (12) derives from the approximate equality $\sqrt{2}\cdot\ln 2=0.9803\ldots\simeq 1$ (if $w_F=0$, then (12) yields $v_{1/2}=w_D/\sqrt{2}\simeq w_D\ \ln 2$). At $A=1/\sqrt{2}$ the width of the Bennett dip exceeds both the power width and the diffusion width by a factor of almost 1.5. Such a contour is depicted in Fig. 2.

Babin *et al.*[5] used a Lorentzian function for an approximate solution. The amplitude and width were found by a variational method. The width of the approximating Lorentzian satisfies the quartic equation

$$v_{1/2}^4-(w^2+w_F^2+w_D^2/2)v_{1/2}^2-2ww_Dv_{1/2}-3w^2w_D^2/2=0, \tag{13}$$

where $w=w_DW$ (in Ref. 5 w was taken to be $w_H=\Gamma_{12}/k$, which is justified if the field is not too strong). Its solution when $w_D,w_F\gg w$ is given by (12), i.e., the halfwidths at half maximum of the solution of (13) and the approximating Lorentzian in the variational method are essentially identical. The level curves for the solution of Eq. (13) in the (w_F^2,w_D^2) plane are segments of straight lines, which become parallel for $w_D,w_F\gg w$, and can thus be mapped into one another by a scale transformation centered at the origin.

The values of $v_{n/4}$, $n=1,2,3$, in Fig. 3 are selected in such a way that the curves $n=1,2,3$ coincide if the dip is a pure Lorentzian. Thus, the distance between curves is the measure of the deviation of the shape from a Lorentzian. The deviation is greater at the center of the dip than in the wings (curve $n=3$ deviates from curve $n=2$ more than does curve $n=1$) and if the power broadening is smaller than the diffusion broadening. When the power width is large, $w_F\gg w_D$, the curves merge, so that saturation effects mask diffusion broadening.

The modified Bessel function $K_\nu(z)$ for half-integer $\nu=n+1/2$, with $n=0,\pm 1,\pm 2,\ldots$, can be expressed in terms of elementary functions. However, the solution (10) of the inhomogeneous equation reduces to elementary functions only if $\alpha=2n+1/2$, $n=0,1,\ldots$. For instance, at $\alpha=1/2$ the contour is exponential (see (11)), while at $\alpha=5/2$ the shape of the dip is given by

$$y(x)=\frac{6-2e^{-|x|}(3+3|x|+x^2)}{x^2},$$

which differs from a Lorentzian of width $\sqrt{13/2}$ (see Eq. (13) and Ref. 5) by 8%.

We now attempt to find the spectrum of a probe field. The shape of the Bennett dip can be found by measuring the absorption spectrum for a probe wave that is in resonance with the same transition as the strong wave. To determine the spectrum of the probe wave, we must augment the system of equations (1), so that it reflects the interaction with the probe wave. In the linear approximation in the probe field intensity, the probe field induces the coherence ρ_{21} and population corrections, which oscillate in time with frequencies Ω_μ and $\Omega_\mu-\Omega$, where Ω_μ is the probe wave offset from resonance. The corrections to the density matrix induced by the probe field are proportional to the population difference when the longitudinal velocity is in resonance with the probe field. If we assume that the rates of excitation of levels 1 and 2 are the same, the population of the long-lived level will be much higher than that of the rapidly decaying. Due to field splitting of levels, the resonant velocity is a nonlinear function of

Ω_μ. This leads to a situation in which the absorption (amplification) spectrum for the probe wave acquires dips whose shape does not follow that of a Bennett dip. The profile is distorted by the nonlinear dependence of the resonant velocity on Ω_μ, that is, by power broadening.

4. SATURATION CURVE

First and foremost, we show that the power absorbed from the field can be expressed solely in terms of the area of the Bennett dip, $\int_{-\infty}^{\infty} dx\ y(x)$. For case 2, substituting $2\,\mathrm{Re}(iG^*r_{21})$ from the first equation in (2), we obtain

$$P(\Omega)=2\hbar\omega\int dv\ \mathrm{Re}(iG^*r_{21})$$
$$=\hbar\omega\Gamma_1\int dv\left[w_{1D}^2\frac{d^2r_1}{dv^2}-r_1+r_1^{(0)}+\frac{A_{21}}{\Gamma_1}(r_2-r_2^{(0)})\right]. \tag{14}$$

Expressing r_1 and r_2 in terms of y, $r_1^{(0)}$, and $r_2^{(0)}$, we have

$$P(\Omega)=\hbar\omega\Gamma_1\int dv(r_1^{(0)}-r_2^{(0)})$$
$$\times\left[y-\frac{d^2y}{dx^2}+\frac{A_{21}}{\Gamma_1}J(1-y)\right].$$

Since the width of the Bennett dip is much less than v_T, we can take $r_1^{(0)}-r_2^{(0)}$ outside the integral sign and evaluate it at $v=\Omega/k$. Noting that $J(1-y)=(y-d^2y/dx^2)\times(1-A_{21}/\Gamma_2)^{-1}$, we obtain

$$P(\Omega)=\frac{\hbar\omega\Gamma_1 w_{1D}(N_1-N_2)}{\sqrt{\pi}v_T(1-A_{21}/\Gamma_2)}$$
$$\times\exp\left[-\left(\frac{\Omega}{kv_T}\right)^2\right]\int_{-\infty}^{\infty}dx\ y(x), \tag{15}$$

where $N_j=\int dv\ r_j^{(0)}$ is the initial concentration of ions in level j. Similarly, for case 1 we have

$$P(\Omega)=\frac{\hbar\omega\Gamma_2 w_{2D}(N_1-N_2)}{\sqrt{\pi}v_T}\exp\left[-\left(\frac{\Omega}{kv_T}\right)^2\right]\int_{-\infty}^{\infty}dx\ y(x). \tag{16}$$

Thus, studying the intensity dependence of the power absorbed from the field reduces to studying the behavior of the integral $\int_{-\infty}^{\infty}dx\ y(x)$, where y is the solution of Eq. (5).

This fact has a simple qualitative explanation. In the steady-state situation, the area of the Bennett dip in level j, i.e., $\int dv(r_j-r_j^{(0)})$, is the ratio of the concentration of ions that absorb (or emit) one field photon per unit time to the population relaxation rate for the given level Γ_j. The power absorbed from the field can be expressed in terms of the area of the Bennett dip if the rate of particle ejection from the level due to collisions is independent of the translational velocity of the particles (e.g., collisions have no effect on the number of particles in the level).

Consider the Fourier transform of Eq. (5):

$$\left(W^2+\frac{A^2}{1+t^2}\right)Y=\frac{d^2Y}{dt^2}+2\pi A^2\delta(t), \tag{17}$$

where

$$Y(t)=(1+t^2)\int_{-\infty}^{\infty}dx\ e^{itx}y(x).$$

The entire dependence of the absorbed power $P(\Omega)$ on the wave's intensity is contained in the amplitude of the zeroth-order harmonic $Y(0)=\int_{-\infty}^{\infty}dx\ y(x)$.

When $W\gg A$, we can neglect the term $A^2/(1+t^2)$ on the left-hand side of Eq. (17), with the result that $Y(t)=\pi A^2\exp(-W|t|)/W$. When $W\gg 1$ or $A\gg 1$, the function $Y(t)$ has a width $(W^2+A^2)^{-1/2}\ll 1$. We then can assume that t is small, $t\ll 1$, and

$$Y(t)=\frac{\pi A^2}{\sqrt{W^2+A^2}}\exp(-\sqrt{W^2+A^2}|t|).$$

In a different limiting case, $W\ll\min\{A,1\}$, we solve Eq. (17) in two ranges t, namely $t\gg 1$ and $t\ll A/W$. When $A\gg W$, these ranges overlap, and when $1\ll t\ll A/W$, both asymptotes are applicable, which makes it possible to match them, with the result that we obtain an approximation for $Y(t)$ that is equally suitable for all values of t.

For $t\gg 1$ we can neglect the 1 in the denominator $1+t^2$, with the result that Eq. (17) reduces to Bessel's differential equation. This yields

$$Y(t)=C_3\sqrt{t}K_\alpha(Wt), \tag{18}$$

where C_3 is a constant determined by matching (18) to the solution $Y(t)$ for small t.

When $t\ll A/W$, we can neglect W^2 on the left-hand side of Eq. (17), whereupon (17) reduces to the hypergeometric equation.

$$Y(t)=C_{1\pm}(i+t)_2F_1\left(\frac{1}{2}+\alpha,\frac{1}{2}-\alpha;2;\frac{i+t}{2i}\right)$$
$$+C_{2\pm}(i-t)_2F_1\left(\frac{1}{2}+\alpha,\frac{1}{2}-\alpha;2;\frac{i-t}{2i}\right); \tag{19}$$

for $t>0$ the constants C_1 and C_2 carry a plus sign, while for $t<0$ they carry a minus sign. Here ${}_2F_1(a,b;c;z)$ is the hypergeometric series (see Ref. 14).

What remains to be done is to find the constants $C_{1\pm}$, $C_{2\pm}$, and C_3. Constraints on the behavior of $Y(t)$ near $t=0$ impose two constraints on the values of the constants; another emerges when we match the asymptotic solutions (18) and (19), and finally, the fact that $y(x)$ is real leads to two more.

Integrating (17) locally near $t=0$, we find that

$$\lim_{t\to+0}Y(t)=\lim_{t\to-0}Y(t),$$
$$\lim_{t\to-0}\frac{dY(t)}{dt}-\lim_{t\to+0}\frac{dY(t)}{dt}=2\pi A^2, \tag{20}$$
$$C_{1+}+C_{2+}=C_{1-}+C_{2-},$$

$$C_{1-}-C_{2-}-C_{1+}+C_{2+}=\frac{2\pi A^2}{{}_2F_1-A^2{}_2F_1^+/4},$$

where

$${}_2F_1={}_2F_1\left(\frac{1}{2}+\alpha,\frac{1}{2}-\alpha;2;\frac{1}{2}\right),$$

$${}_2F_1^+={}_2F_1\left(\frac{3}{2}+\alpha,\frac{3}{2}-\alpha;3;\frac{1}{2}\right).$$

Since $y(x)$ is real, $Y(-t)=Y^*(t)$, so that $C_{m-}=-C_{m+}^*$, with $m=1,2$.

When $1\ll t\ll A/W$, the solutions (18) and (19) must coincide. Here the argument of the hypergeometric series in (19) is large, $t\gg 1$, and in (18) the argument of the modified Bessel function can be chosen to be small, $Wt\ll 1$. Expanding (18) as a power series in t at $t=0$ (see Ref. 13) and (19) as a power series in t^{-1} at $t=\infty$ (see Ref. 14),

$$\begin{aligned}
Y(t)&=\frac{\pi C_3\sqrt{t}}{2\sin\pi\alpha}\left[\frac{(Wt/2)^{-\alpha}}{\Gamma(1-\alpha)}-\frac{(Wt/2)^{\alpha}}{\Gamma(1+\alpha)}\right]+\cdots,\\
Y(t)&=\exp\left[-\frac{i\pi}{2}\left(\alpha+\frac{1}{2}\right)\right](C_{1+}-ie^{i\pi\alpha}C_{2+})\\
&\quad\times\frac{\Gamma(-2\alpha)}{\Gamma(1/2-\alpha)\Gamma(3/2-\alpha)}\left(\frac{t}{2}\right)^{1/2-\alpha}\\
&\quad+\exp\left[\frac{i\pi}{2}\left(\alpha-\frac{1}{2}\right)\right](C_{1+}-ie^{-i\pi\alpha}C_{2+})\\
&\quad\times\frac{\Gamma(2\alpha)}{\Gamma(1/2+\alpha)\Gamma(3/2+\alpha)}\left(\frac{t}{2}\right)^{1/2+\alpha}+\cdots,
\end{aligned}\tag{21}$$

and equating the coefficients of $t^{1/2\pm\alpha}$, we obtain

$$\begin{aligned}
&\frac{C_{1+}-ie^{-i\pi\alpha}C_{2+}}{C_{1+}-ie^{i\pi\alpha}C_{2+}}=\varepsilon e^{-i\pi\alpha},\\
&\varepsilon=-W^{2\alpha}\frac{\Gamma(-2\alpha)\Gamma(1-\alpha)\Gamma(1/2+\alpha)\Gamma(3/2+\alpha)}{\Gamma(2\alpha)\Gamma(1+\alpha)\Gamma(1/2-\alpha)\Gamma(3/2-\alpha)}.
\end{aligned}\tag{22}$$

Representing C_{1+} in the form $C_{1+}=\mu C_{2+}$, we obtain

$$\mu=ie^{-i\pi\alpha}\frac{1-\varepsilon e^{i\pi\alpha}}{1-\varepsilon e^{-i\pi\alpha}},\quad |\mu|=1,$$

$$\begin{aligned}
Y(0)&=iC_{2+}{}_2F_1=-\frac{\pi A^2{}_2F_1}{2({}_2F_1-A^2{}_2F_1^+/4)}\frac{|1+\mu|^2}{\operatorname{Im}\mu}\\
&=\frac{\pi A^2{}_2F_1}{{}_2F_1-A^2{}_2F_1^+/4}\cot\left(\frac{\pi}{2}\left(\alpha-\frac{1}{2}\right)\right.\\
&\quad\left.+\arctan\frac{\varepsilon\sin\pi\alpha}{1-\varepsilon\cos\pi\alpha}\right).
\end{aligned}\tag{23}$$

Thus, when the power width exceeds the radiative width ($A\gg W$), the power absorbed from the field per unit volume is given by (16) for case 1 and by (15) for case 2, with $\int_{-\infty}^{\infty}dx\ y(x)=Y(0)$ given by (23). The power width is comparable to the diffusion width, $A\sim 1$. The expression (23) is cumbersome, so we start with several limiting cases, after which we suggest a simple but accurate approximation.

4.1. Limiting cases

Suppose that the field is weak ($A\ll 1$). Then $\alpha\simeq 1/2+A^2$, $\varepsilon\simeq W^{2\alpha}$, and the argument of $\cot z$ in (23) is small, with the result that

$$Y(0)\simeq\frac{\pi A^2}{W^{2\alpha}+\pi A^2/2}.\tag{24}$$

A similar result, $Y(0)=\pi A^2/(W+\pi A^2/2)$, which differs from (24) only in the exponent of W in the denominator, can be obtained by replacing $A^2/(W^2+x^2)$ in (5) with $\pi A^2\delta(x)/W$ (Ref. 7). The extent to which the approximation based on this substitution differs from (24) can be estimated by the difference $1-W^{2A^2}$, which for $W\ll A\ll 1$ can be of order unity. The expression (24) describes so-called *homogeneous saturation*, where the diffusion shape of the Bennett dip does not change (it remains exponential), while the depth of the dip, $y(0)$, reaches its maximum value, equal to unity, as the field strength increases.

Now consider strong fields ($A\sim 1$). We can then neglect the correction to the argument of $\cot z$ in (23), which is related to the fact that ε is nonvanishing (see Appendix). Setting $\varepsilon=0$ in (23), we can write the asymptotic formulas for $Y(0)$ for strong and weak diffusion broadening as compared to power broadening (see (11)):

$$Y(0)=\begin{cases}2, & A\ll 1,\\ \pi A(1+1/4\,A^2), & A\gg 1.\end{cases}\tag{25}$$

4.2. Interpolation formulas

For strong fields ($A\sim 1$), $Y(0)$ obeys the interpolation formula

$$Y(0)\simeq\pi\sqrt{(2/\pi)^2+A^2},\tag{26}$$

which, as numerical calculations show, is valid in the region where $W\ll\min\{A^2,1\}$ with an accuracy no worse than 2%.

Combining (26) and (24), we arrive at an interpolation formula with an applicability range broader than (26):

$$Y(0)\simeq\frac{\pi A^2\sqrt{(2/\pi)^2+A^2}}{(2/\pi)W^{2\alpha}+A^2}.\tag{27}$$

This expression is valid in the region $W\ll 1$ and, in particular, in the case of a weak field, $A^2\sim W$.

Finally, altering (27) somewhat, we obtain at the interpolation formula

$$Y(0)\simeq\frac{\pi A^2\sqrt{(2/\pi)^2+W^2+A^2}}{(2/\pi)W^{2\alpha}/(1+2W^{2\alpha})+W^2+A^2},\tag{28}$$

which is valid for all A, $W>0$ with an accuracy of about 3%. Examples of saturation curves (which reflect the dependence of the power absorbed from the field on the field intensity) calculated with (28) are depicted in Fig. 4. We see that the smaller the natural width W, the sooner homogeneous saturation sets in (the transition from curve *1* to curves *2* and *3* in Fig. 4b). However, a further increase in A^2 (Fig. 4a) leads to a square-root increase in the absorbed power. This increase is almost unnoticeable on the scale of Fig. 4b.

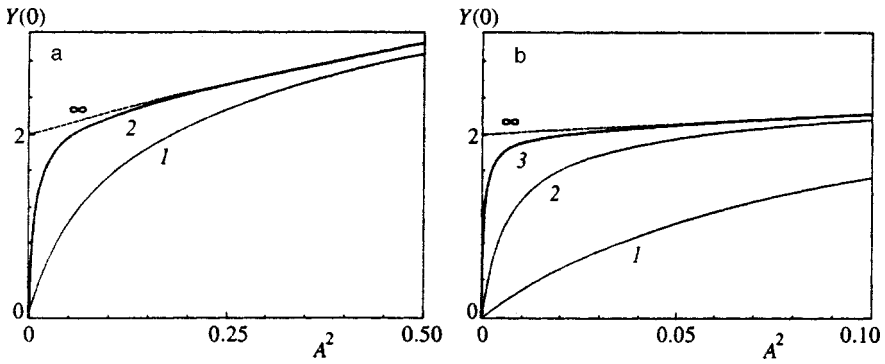

FIG. 4. The absorbed power $Y(0)$ as a function of the field intensity A^2. Curve n corresponds to the value $W=10^{-n}$ (curve ∞ corresponds to $W=0$).

Reverting to the original variables of the problem, we obtain

$$P(\Omega)=\frac{2\sqrt{\pi}\hbar\omega\Gamma_{12}|G|^2(N_1-N_2)}{kv_T}\exp\left[-\left(\frac{\Omega}{kv_T}\right)^2\right]\times\frac{\sqrt{\Gamma_{12}^2+\mathscr{F}^2+(2/\pi)^2\mathscr{D}}}{\Gamma_{12}^2+\mathscr{F}^2+(2/\pi)\mathscr{D}p/(1+2p)},$$

$$\mathscr{F}^2=2\Gamma_{12}|G|^2\left(\frac{1}{\Gamma_1}+\frac{1}{\Gamma_2}-\frac{A_{21}}{\Gamma_1\Gamma_2}\right),\quad \mathscr{D}=\frac{\nu k^2 v_T^2}{2\Gamma_<},$$

$$p=\left(\frac{\Gamma_{12}^2+2\Gamma_{12}|G|^2/\Gamma_>}{\mathscr{D}}\right)^{\sqrt{\mathscr{F}^2/\mathscr{D}+1/4}},$$

where $\Gamma_<$ ($\Gamma_>$) is the lesser (greater) of the two quantities Γ_1 and Γ_2. In case 1 we have $\Gamma_<=\Gamma_2$ and $\Gamma_>=\Gamma_1$, while in case 2 we have $\Gamma_<=\Gamma_1$ and $\Gamma_>=\Gamma_2$.

5. DISCUSSION

Two results have been obtained in the previous sections.

1. We derived an expression for the shape of the Bennett dip in a two-level system with a rapidly decaying level, for an arbitrary ratio of the power and diffusion widths of the dip. The expression is valid if the natural width of the dip is small compared to the total width and the width of the Maxwellian distribution exceeds that of the dip. We show that the square of the total dip width at half maximum is the sum of squares of the power and diffusion widths.

2. Under the above conditions, we derived a formula that reflects the dependence of the absorbed power on the intensity of the incident wave. The formula describes the well-known limiting cases of low and high intensity and the smooth transition from homogeneous saturation to inhomogeneous saturation in a new intermediate region. An interpolation formula has been suggested for this dependence. The formula is valid even if the homogeneous width is not small.

5.1. Shape of dip

Equation (5) contains three widths: the diffusion width (equal to unity), the power width A, and the homogeneous width W. The solutions for the shape of the Bennett dip in the collisionless case (both A and W are large compared to unity) and in the case of a weak field (A is small compared to W and unity) were obtained earlier. The only case that remained unexplored was when one of the three widths is small and the other two are comparable (W small). Here we add power broadening to diffusion broadening. The depth $y(0)$ of the Bennett dip for $A^2\gg W$ reaches its maximum $y(0)=1$, while the shape of the dip for $A\ll 1$ is determined by diffusion and is described by an exponential contour. As the field intensity A^2 increases, the dip flattens out. The general case, in which all three widths are comparable, can hardly be solved in terms of hypergeometric functions, which usually contain only two widths. For instance, $J_\nu(z)$ contains two scales: unity (the period of oscillations at large values of z), and ν (the distance from $z=0$ to the origin of oscillations). However, even if we were able to write the solution in some form, it would still be simpler to find it by direct numerical methods in each specific case. Furthermore, when W is of order unity the shape of the dip is close to Lorentzian, so that the variational approximation described in Ref. 5 yields satisfactory results.

5.2. Saturation curve

The expression for the absorbed power obtained in Ref. 7 and formula (9) of the present paper can be applied in two regions: $A^2\ll 1/|\ln W|$, $W\ll 1$, and $W\ll A^2\ll 1$. This provides a correct description of homogeneous saturation for $W\ll 1$. The condition that they can be used to describe homogeneous saturation is the smallness of the parameter $W|\ln W|$, which tends to zero as $W\to 0$. Formula (9) cannot be applied in the case of strong fields ($A\sim 1$), when the power width of the Bennett dip is of the order of the diffusion width. One property of (9) is that for large values of A, the integral $\int dx\, y(x)$ reaches its maximum value, equal to 2. It was unclear how the absorbed power behaves at $A\sim 1$, since it is known that $\int dx y(x)=\pi A$ for $A\gg 1$, i.e., the absorbed power grows with the wave intensity. The expression (23) or the interpolation formulas (27) and (28) describe precisely the case where $A\sim 1$. They can also be applied for weak fields ($A\ll 1$). These expressions cannot be applied when the width of one of the levels is small compared to that of the other, or when the width of the Bennett dip is small compared to the width v_T of the Maxwellian distribution. If the diffusion width of the Bennett dip is less than v_T, saturation associated with the fact that the power width becomes equal to v_T (and hence the square-root increase in the absorbed power, $\int dx\, y(x)=\pi\sqrt{A^2}$, with the field intensity A^2 becoming limited to the extent that the absorbed power finally becomes a constant) can be described without taking diffusion into ac-

count ($A \gg 1$). This limiting case has been thoroughly discussed in the literature (see Ref. 15). When saturation is homogeneous, the increase in the absorbed power with the field intensity slows down as the depth $y(0)$ of the dip approaches unity. However, in strong fields the dip flattens out, and because of this, the increase in the absorbed power does not stop. The transition from the homogeneous saturation region to the dip-flattening regime occurs at $A \sim W^{1/4}$, when the increase in the absorbed power due to homogeneous saturation slows down so much that it becomes comparable to the increase due to dip flattening.

We are grateful to E. V. Podivilov for fruitful discussions and to A. I. Chernykh for consultations on numerical methods. The present work was partially supported by the Russian Fund for Fundamental Research (Grant Nos. 96-02-19052 and 96-15-96642), the Interdisciplinary Science and Technology Program in Optics and Laser Physics (Grant 1.53), and the Soros Students Program (M.G.S, Grant s97-215).

APPENDIX

The correction to the argument of $\cot z$ in (23) related to the finite value of ε can become important in three cases:

1. ε is large (α is close to 1,2, . . . ,n, . . .).

2. The value of $\cot z$ is large (α is close to 5/2,9/2, . . . , $2n+1/2$, . . .).

3. The value of $\cot z$ is small (α is close to 3/2,7/2, . . . , $2n-1/2$, . . .).

Let $\alpha = n + \delta$, with $\delta \ll 1$. Then $\varepsilon \simeq (-1)^{n+1} C_n W^{2\alpha}/\delta^2$, with C_n positive (e.g., $C_1 = 3/64$), $\sin \pi\alpha \simeq (-1)^n \delta$, and $\cos \pi\alpha \simeq (-1)^n$. The argument of arctan z in (23) can be written as

$$\frac{\varepsilon \sin \pi\alpha}{1-\varepsilon \cos \pi\alpha} \simeq -\frac{\pi\delta C_n W^{2\alpha}}{C_n W^{2\alpha}+\delta^2}.$$

It is at its maximum at $\delta \sim W^\alpha$; the maximum value of the argument is of order $W^\alpha \ll 1$.

Let $\alpha = 2n + 1/2 + \delta$, with $\delta \ll 1$. Then $\varepsilon \simeq C_n W^{2\alpha}\delta$ (e.g., $C_1 = 2/675$), $\sin \pi\alpha \simeq 1$, and $\cos \pi\alpha \simeq 0$. The value of $\cot z$ in (23) is given by the approximate expression

$$\cot z \simeq \frac{1}{(\pi/2 + C_n W^{2\alpha})\delta},$$

i.e., the correction related to $\varepsilon \neq 0$ is small ($W \ll 1$). Note that if we put $n = 0$ (a weak field), then $\varepsilon \simeq W^{2\alpha}$, i.e., ε is not small in terms of δ.

Finally, let $\alpha = 2n - 1/2 + \delta$, with $\delta \ll 1$. Then $\varepsilon \simeq -C_n W^{2\alpha}\delta$ (e.g., $C_1 = 2/9$), $\sin \pi\alpha \simeq -1$, and $\cos \pi\alpha \simeq 0$. The value of $\cot z$ in (23) is given by the approximate expression

$$\cot z \simeq -(\pi/2 + C_n W^{2\alpha})\delta,$$

i.e., the correction related to $\varepsilon \neq 0$ is small ($W \ll 1$).

*)E-mail: 249@okibox.iae.nsk.su

[1] W. R. Bennett, Jr., Phys. Rev. **126**, 580 (1962).
[2] L. D. Landau, *Collected Papers* [in Russian], Nauka, Moscow (1969), p. 199.
[3] G. I. Smirnov and D. A. Shapiro, Zh. Éksp. Teor. Fiz. **76**, 2084 (1979) [Sov. Phys. JETP **49**, 1054 (1979)].
[4] R. Karplus and J. Schwinger, Phys. Rev. **73**, 1020 (1948).
[5] S. A. Babin, V. I. Donin, and D. A. Shapiro, Zh. Éksp. Teor. Fiz. **91**, 1270 (1986) [Sov. Phys. JETP **64**, 750 (1986)].
[6] D. A. Shapiro and M. G. Stepanov, J. Phys. B **30**, L377 (1997).
[7] L. B. Kurlaev and D. A. Shapiro, Kvant. Élektron. (Moscow) **21**, 1080 (1994) [Quantum Electron. **24**, 1003 (1994)].
[8] A. A. Apolonsky, S. A. Babin, A. I. Chernykh, S. I. Kablukov, S. V. Khorev, E. V. Podivilov, and D. A. Shapiro, Phys. Rev. A **55**, 661 (1997).
[9] S. A. Babin and D. A. Shapiro, Phys. Rep. **241**, 119 (1994).
[10] E. V. Podivilov, D. A. Shapiro, and M. G. Stepanov, Phys. Rev. Lett. **74**, 3979 (1995); Zh. Éksp. Teor. Fiz. **109**, 418 (1996) [JETP **82**, 221 (1996)].
[11] S. G. Rautian, Zh. Éksp. Teor. Fiz. **51**, 1176 (1966) [Sov. Phys. JETP **24**, 788 (1967)].
[12] A. Erdélyi, *Higher Transcendental Functions*, (Bateman Project) Vol. 3, McGraw-Hill, New York (1955); I. V. Komarov, L. I. Ponomarev, and S. Yu. Slavyanov, *Spheroidal and Coulomb Spheroidal Functions* [in Russian], Nauka, Moscow (1976).
[13] A. Erdélyi, *Higher Transcendental Functions*, (Bateman Project) Vol. 2, McGraw-Hill, New York, (1953).
[14] A. Erdélyi, *Higher Transcendental Functions*, (Bateman Project) Vol. 1, McGraw-Hill, New York (1953).
[15] S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in the Spectra of Atoms and Molecules* [in Russian], Nauka, Novosibirsk (1979).

Translated by Eugene Yankovsky

Dependence of transport characteristics of molecules on the rotational quantum number

A. I. Parkhomenko and A. M. Shalagin*)

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

(Submitted 29 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 1649–1660 (May 1998)

We derive a simple analytic formula that describes the relative difference of transport collision rates, $\Delta\nu/\nu$, for collisions of molecules and atoms in the rovibrational excitation of the former by light, as a function of the rotational quantum numbers of the combining (i.e., affected by radiation) levels of the molecules. (The relative difference of transport collision rates can be measured in light-induced drift, or LID, experiments and is proportional to the LID effect.) The formula is valid in the energy sudden approximation and is based on the well-known factorization formula for cross sections of RT-transitions in linear molecules that collide with atoms. We show that in this approximation the factor $\Delta\nu/\nu$ is the sum of two independent terms, the vibrational term $(\Delta\nu/\nu)_{\rm vib}$ and the rotational term $(\Delta\nu/\nu)_{\rm rot}$. Each term can be measured individually in LID experiments. © *1998 American Institute of Physics.* [S1063-7761(98)00705-7]

1. INTRODUCTION

Light-induced drift, or LID,[1] predicted in Ref. 2, is being actively investigated both theoretically and experimentally.[3–6] This effect consists in the formation of a macroscopic flux of absorbing particles that interact with a traveling light wave and collide with buffer gas particles. The magnitude of the LID effect is proportional to $\Delta\nu/\nu \equiv(\nu_m-\nu_n)/\nu_n$, the relative difference of the transport collision rates ν_α $(\alpha=m,n)$ of the collision of resonant particles and buffer particles in the ground (n) and excited (m) states. This is the basis of one of the fundamental scientific applications of LID, namely measuring the relative variations of the transport collision rates caused by excitation of the particles.

Recent LID experiments involving molecules[7–9] have revealed unexpectedly strong dependence (which remains unexplained) of the factor $\Delta\nu/\nu$ on the rotational quantum numbers of the molecules.

A theoretical explanation of this dependence would make it possible to predict the magnitude of the LID effect for specific molecules and to calculate the variations in the transport collision rates of molecules under rovibrational excitation. These results would certainly be pertinent to the separation of mixtures and isotopes of molecular gases using LID. They would also be of interest for the physics of intermolecular interactions.

In this paper we solve for the dependence of the factor $\Delta\nu/\nu$ on the rotational numbers for linear molecules in the sudden approximation, which is effective for molecules with a moderate value of the rotational constant.

2. GENERAL RELATIONSHIPS FOR COLLISION RATES

We consider LID in the field of a traveling light wave. The interaction between the radiation and the molecules in the buffer gas are described by the following transport equations:

$$\frac{d}{dt}\rho_m(J_m,\mathbf{v})=S_m(J_m,\mathbf{v})+NP(\mathbf{v})\,\delta_{J_mJ_{m0}},$$
$$\frac{d}{dt}\rho_n(J_n,\mathbf{v})=S_n(J_n,\mathbf{v})-NP(\mathbf{v})\,\delta_{J_nJ_{n0}}. \tag{1}$$

Here $\rho_\alpha(J_\alpha,\mathbf{v})$ is the population distribution for the absorbing molecules over velocity $\mathbf{v}$ and the rotational levels J_α in the vibrational state α ($\alpha=n$ is the vibrational ground state, $\alpha=m$ is a vibrational excited state, and J_α denotes the set of rotational quantum numbers characterizing the rotational state), $S_\alpha(J_\alpha,\mathbf{v})$ is the collision integral reflecting the collisions of buffer particles and molecules in the vibrational state α and the rotational state J_α, $P(\mathbf{v})$ is the probability (per unit time) that a molecule with fixed velocity $\mathbf{v}$ absorbs radiation, and $N=N_m+N_n$ is the concentration of the absorbing molecules, with

$$N_\alpha=\sum_{J_\alpha}\int\rho_\alpha(J_\alpha,\mathbf{v})d\mathbf{v}.$$

In Eqs. (1) we assume that the radiation is in resonance with the rovibrational transition nJ_{n0}–mJ_{m0}. Here we ignore radiative relaxation, since it affects rovibrational transitions only at very low pressures.

An absorbing molecules as a whole is subject to a friction force

$$\mathbf{F}=\mathbf{F}_m+\mathbf{F}_n, \quad \mathbf{F}_\alpha=M\sum_{J_\alpha}\int \mathbf{v}S_\alpha(J_\alpha,\mathbf{v})d\mathbf{v},$$

$$\alpha=m,n, \tag{2}$$

due to collisions of excited $\alpha=m$ and unexcited $\alpha=n$ molecules and particles of the buffer gas (M is the mass of a molecule). Obviously, the partial frictional forces $\mathbf{F}_\alpha$ oppose the partial fluxes $\mathbf{j}_\alpha=\Sigma_{J_\alpha}\int \mathbf{v}\rho_\alpha(J_\alpha,\mathbf{v})d\mathbf{v}$ of molecules in states α:

$$\mathbf{F}_\alpha=-M\nu_\alpha\mathbf{j}_\alpha, \quad \alpha=m,n. \tag{3}$$

The proportionality factor ν_α has both the dimensions and sense of a collision rate. In steady-state and spatially homogeneous conditions, the initial equations (1) and the definition of $\mathbf{F}_\alpha$ in (2) imply $\mathbf{F}=0$ (steady-state flow). Using this fact and Eqs. (2) and (3), we find the following expression for the total absorbing-particle flux $\mathbf{J}=\mathbf{j}_m+\mathbf{j}_n$ (see Ref. 10):

$$\mathbf{J}\equiv N\mathbf{u}_0=\frac{\nu_n-\nu_m}{\nu_n}\mathbf{j}_m, \tag{4}$$

where $\mathbf{u}_0$ is the LID velocity. In steady-state and spatially homogeneous conditions we find that multiplying the first equation in (1) by $\mathbf{v}$, summing the product over J_m, and allowing for (2) and (3) yields

$$\mathbf{j}_m=\frac{N}{\nu_m}\int \mathbf{v}P(\mathbf{v})d\mathbf{v}. \tag{5}$$

As a result, for the LID velocity we have the well-known formula[11]

$$\mathbf{u}_0=\frac{\nu_n-\nu_m}{\nu_n\nu_m}\int \mathbf{v}P(\mathbf{v})d\mathbf{v}. \tag{6}$$

This formula has been used to process the results of most experiments in LID of molecules. In particular, having the experimental data, we can use this formula to calculate the factor $\Delta\nu/\nu\equiv(\nu_m-\nu_n)/\nu_n$, which is the relative difference of the transport collision rates.

We now establish the import of the collision rates ν_m and ν_n in (6), so that we can see how they depend on the rotational quantum number. The collision integral in Eqs. (1) and (2) has the structure

$$S_\alpha(J_\alpha,\mathbf{v})=\sum_{J_{\alpha1}}\int [A(\alpha J_{\alpha1}\mathbf{v}_1\to\alpha J_\alpha\mathbf{v})\rho_\alpha(J_{\alpha1},\mathbf{v}_1) - A(\alpha J_\alpha\mathbf{v}\to\alpha J_{\alpha1}\mathbf{v}_1)\rho_\alpha(J_\alpha,\mathbf{v})]d\mathbf{v}_1. \tag{7}$$

We note that the two kernels of the collision integral, $A(\alpha J_{\alpha1}\mathbf{v}_1\to\alpha J_\alpha\mathbf{v})$ and $A(\alpha J_\alpha\mathbf{v}\to\alpha J_{\alpha1}\mathbf{v}_1)$, describe inelastic ($J_\alpha\neq J_{\alpha1}$) collisional transitions between the rotational states of the given vibrational level α and elastic ($J_\alpha=J_{\alpha1}$) collisions in the rotational state J_α. In (7) the collisional transitions $m\to n$ and $n\to m$ between the vibrational levels are ignored. Using (7), we can transform the frictional force $\mathbf{F}_\alpha$ in (2) to

$$\mathbf{F}_\alpha=M\sum_{J_\alpha}\sum_{J_{\alpha1}}\int (\mathbf{v}_1-\mathbf{v})\rho_\alpha(J_\alpha,\mathbf{v}) \times A(\alpha J_\alpha\mathbf{v}\to\alpha J_{\alpha1}\mathbf{v}_1)d\mathbf{v}\,d\mathbf{v}_1. \tag{8}$$

The kernel of the collision integral is given by the following formula:[5,12]

$$A(\alpha J_\alpha\mathbf{v}\to\alpha J_{\alpha1}\mathbf{v}_1)=2\int \rho_b(\mathbf{v}-\mathbf{u}) \times|f(\alpha J_\alpha\mathbf{u}\to\alpha J_{\alpha1}\mathbf{u}_1)|^2 \times\delta\left(\mathbf{v}_1-\mathbf{v}-\frac{\mu}{M}(\mathbf{u}_1-\mathbf{u})\right) \times\delta\left(\mathbf{u}_1^2-\mathbf{u}^2+\frac{2\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)}{\mu}\right)d\mathbf{u}\,d\mathbf{u}_1, \tag{9}$$

Here $\mathbf{u}$ and $\mathbf{u}_1$ are the relative velocities of the colliding particles before and after collision, $\rho_b(\mathbf{v}-\mathbf{u})$ is the velocity distribution of the buffer particles (which are assumed structureless), $f(\alpha J_\alpha\mathbf{u}\to\alpha J_{\alpha1}\mathbf{u}_1)$ is the scattering amplitude in the $J_\alpha\to J_{\alpha1}$ channel with the relative velocity changing from $\mathbf{u}$ to $\mathbf{u}_1$, μ is the reduced mass of the colliding particles, $\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)=\varepsilon_\alpha(J_{\alpha1})-\varepsilon_\alpha(J_\alpha)$ is the change in rotational energy of the molecules due to inelastic transitions $J_\alpha\to J_{\alpha1}$ in the vibrational state α, and $\varepsilon_\alpha(J_\alpha)$ is the rotational energy of state J_α of rotational level α. Substituting (9) into (8), we obtain the following expression for the partial frictional force $\mathbf{F}_\alpha$ (see Ref. 5):

$$\mathbf{F}_\alpha=-M\sum_{J_\alpha}\int \mathbf{v}\rho_\alpha(J_\alpha,\mathbf{v})\nu_\alpha(\mathbf{v},J_\alpha)d\mathbf{v}. \tag{10}$$

Here we have introduced the transport collision rates $\nu_\alpha(\mathbf{v},J_\alpha)$ described by

$$\nu_\alpha(\mathbf{v},J_\alpha)=\frac{\mu}{M}\sum_{J_{\alpha1}}\int u\,\frac{\mathbf{u}\cdot\mathbf{v}}{v^2}\rho_b(\mathbf{v}-\mathbf{u}) \times\sigma_\alpha^{\mathrm{tr}}(u,\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha);\ J_\alpha\to J_{\alpha1})d\mathbf{u},$$

$$\sigma_\alpha^{\mathrm{tr}}(u,\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha);\ J_\alpha\to J_{\alpha1})=\int\left(1-\frac{\mathbf{u}\cdot\mathbf{u}_{1J_\alpha J_{\alpha1}}}{u^2}\right) \times\sigma_\alpha(J_\alpha,\mathbf{u}\to J_{\alpha1},\mathbf{u}_1)d\mathbf{n}_1, \tag{11}$$

$$u_{1J_\alpha J_{\alpha1}}=\sqrt{u^2-\frac{2\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)}{\mu}}, \quad \mathbf{n}_1\equiv\frac{\mathbf{u}_1}{u_1},$$

$$\sigma_\alpha(J_\alpha,\mathbf{u}\to J_{\alpha1},\mathbf{u}_1)=\frac{u_{1J_\alpha J_{\alpha1}}}{u}|f(\alpha J_\alpha\mathbf{u}\to\alpha J_{\alpha1}\mathbf{u}_1)|^2,$$

and $\sigma_\alpha(J_\alpha,\mathbf{u}\to J_{\alpha1},\mathbf{u}_1)$ and $\sigma_\alpha^{\mathrm{tr}}(u,\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha);J_\alpha\to J_{\alpha1})$ are the differential and transport scattering cross section, respectively.

If the transport collision rates $\nu_\alpha(\mathbf{v},J_\alpha)$ in Eq. (10) are independent of velocity and rotational number, we obtain Eq. (3) and, as a result, Eq. (6), where ν_m and ν_n are completely determined by the properties of the medium and are independent of the parameters of the laser light (frequencies, intensities, and types of excited transition). Generally, however, ν_m and ν_n depend on the specific distribution over velocities and rotational levels. It would seem that to find ν_m and ν_n we might again have to revert to solving the original transport

equations. However, in real situations the transport collision rates change very gradually over the velocity and rotational-number intervals important for the process. This makes it possible to either neglect such changes or approximately account for them with satisfactory accuracy.

3. APPROXIMATE EXPRESSIONS FOR TRANSPORT CHARACTERISTICS

The results of many LID experiments involving molecules (see, e.g., Refs. 13–15 and the literature cited therein) show that the model of velocity-independent transport collision rates usually yields satisfactory results. Only in rare cases in which the difference between ν_m and ν_n is extremely small does the velocity dependence of the transport rates become significant and lead to so-called anomalous LID.[7,15–17] In this paper we focus on the J-dependence of LID, which is related, as we have just seen, to the J-dependence of the transport collision rates $\nu_\alpha(\mathbf{v},J_\alpha)$. As for the v-dependence of $\nu_\alpha(\mathbf{v},J_\alpha)$, we employ the model of velocity-independent transport collision rates. To this end in (10) we make the substitution

$$\nu_\alpha(\mathbf{v};J_\alpha)\to\nu_\alpha(J_\alpha). \tag{12}$$

For the sake of definiteness we take $\nu_\alpha(J_\alpha)$ in the form

$$\nu_\alpha(J_\alpha)=\frac{M}{kT}\int(\mathbf{n}\cdot\mathbf{v})^2W(\mathbf{v})\,\nu_\alpha(\mathbf{v},J_\alpha)d\mathbf{v}, \tag{13}$$

where $\mathbf{n}$ is a randomly directed unit vector, and $W(\mathbf{v})$ is the equilibrium Maxwellian distribution for the absorbing particles. Clearly, if $\nu_\alpha(\mathbf{v},J_\alpha)$ in (13) is velocity-independent, this equation becomes an identity (if we allow for (12)). Our choice of $\nu_\alpha(J_\alpha)$ in the form (13) is convenient because (13) is the usual relation for introducing the average transport rate ν, which is simply related to the diffusion coefficient:

$$D=\frac{kT}{M\nu}. \tag{14}$$

If we allow for the substitution (12), then using Eq. (10) and Eq. (6) we obtain the following expression for ν_α ($\alpha=m,n$):

$$\nu_\alpha=\frac{\Sigma_{J_\alpha}\nu_\alpha(J_\alpha)j_\alpha(J_\alpha)}{\Sigma_{J_\alpha}j_\alpha(J_\alpha)}, \tag{15}$$

where j_α is the partial flux of absorbing particles in rotational level J_α of the vibrational state α.

Laser light induces a flux $j_\alpha(J_{\alpha0})$ of particles in rotational level $J_{\alpha0}$. This flux is partially transferred to neighboring rotational levels J_α by collisions. In the process, the flux slows down. The effective interval of rotational levels in the neighborhood of $J_{\alpha0}$ where $j_\alpha(J_\alpha)$ differs significantly from zero can be assumed small compared to the interval in which variations of $\nu_\alpha(J_\alpha)$ are significant. On the basis of this, we can take $j_\alpha(J_\alpha)$ with the value $J_\alpha=J_{\alpha0}$ outside the sum in (15), i.e., at the point where the flux $j_\alpha(J_\alpha)$ is at its maximum. As a result, Eq. (15) yields

$$\nu_\alpha=\nu_\alpha(J_{\alpha0}). \tag{16}$$

Thus, in the given approximation the J-dependence of LID is explicitly related to the J-dependence of the transport collision rates.

On the basis of (13) and (11) we can derive the following expression for the transport rates $\nu_\alpha(J_\alpha)$:

$$\nu_\alpha(J_\alpha)=\frac{2a}{3(kT)^3}\sum_{J_{\alpha1}}\int_{f(\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha))}^{\infty}dE\;E^2\exp\left(-\frac{E}{kT}\right)\times\int d\Omega\left(1-\sqrt{1-\frac{\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)}{E}}\cos\,\theta\right)\times\sigma_\alpha(E,\theta,\phi;\;J_\alpha\to J_{\alpha1}), \tag{17}$$

where

$$a=\frac{\mu}{M}N_bv_T,\quad v_T=\sqrt{\frac{8kT}{\pi\mu}},$$

$$E=\frac{\mu u^2}{2},\quad \cos\,\theta=\frac{\mathbf{u}\cdot\mathbf{u}_{1J_\alpha J_{\alpha1}}}{uu_{1J_\alpha J_{\alpha1}}},$$

$$f(x)=\begin{cases}0, & x\leqslant0,\\ x, & x>0,\end{cases}\quad d\Omega=\sin\,\theta\,d\theta\,d\phi,$$

θ and ϕ are the polar and azimuthal scattering angles, $\sigma_\alpha(E,\theta,\phi;J_\alpha\to J_{\alpha1})$ is the differential scattering cross section in the channel $J_\alpha\to J_{\alpha1}$, and E is the kinetic energy of the relative motion of the colliding particles. Note that in the integral with respect to E in (17) the lower integration limit, $f(\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha))$, can formally be set to zero, since for $\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)>0$ in the energy range $0\leqslant E<\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)$ we have $\sigma_\alpha(E,\theta,\phi;J_\alpha\to J_{\alpha1})=0$ due to energy conservation.

For subsequent analysis it is convenient to transform the integral for $\nu_\alpha(J_\alpha)$ in (17) and represent it as a sum of two terms,

$$\nu_\alpha(J_\alpha)=\nu_\alpha^t(J_\alpha)+\nu_\alpha^t(J_\alpha), \tag{18}$$

where

$$\nu_\alpha^t(J_\alpha)=\frac{2a}{3(kT)^3}\int_0^\infty E^2\,\exp\left(-\frac{E}{kT}\right)\sigma_\alpha^t(E,J_\alpha)\;dE, \tag{19}$$

$$\nu_\alpha^c(J_\alpha)=\frac{2a}{3(kT)^3}\sum_{J_{\alpha1}}\int_0^\infty dE\;E^2\;\exp\left(-\frac{E}{kT}\right)\times\left(1-\sqrt{1-\frac{\Delta\varepsilon_\alpha(J_{\alpha1}J_\alpha)}{E}}\right)\sigma_\alpha^c(E;J_\alpha\to J_{\alpha1}). \tag{20}$$

Here we have introduced the notation

$$\sigma_\alpha^t(E,J_\alpha)=\int(1-\cos\,\theta)\left(\sum_{J_{\alpha1}}\sigma_\alpha(E,\theta,\phi;\;J_\alpha\to J_{\alpha1})\right)d\Omega, \tag{21}$$

$$\sigma_\alpha^c(E;J_\alpha\to J_{\alpha1})=\int\cos\,\theta\sigma_\alpha(E,\theta,\phi;\;J_\alpha\to J_{\alpha1})d\Omega.$$

For subsequent calculations we use the well-known formula that relates the differential cross sections of RT-transitions in linear molecules when they collide with

atoms,[18–21] which is derived in the sudden-perturbation approximation commonly used in the theory of inelastic molecular collisions

$$\sigma_\alpha(E,\theta,\phi;\ J_\alpha\to J_{\alpha 1})=\frac{E+\varepsilon_\alpha(J_\alpha)}{E}(2J_{\alpha 1}+1)\times\sum_{L=|J_\alpha-J_{\alpha 1}|}^{L=J_\alpha+J_{\alpha 1}}\begin{pmatrix}J_\alpha & J_{\alpha 1} & L\\ 0 & 0 & 0\end{pmatrix}^2\times\sigma_\alpha(E+\varepsilon_\alpha(J_\alpha),\theta,\phi;\ 0\to L),\tag{22}$$

where $\begin{pmatrix}abc\\000\end{pmatrix}$ is the Wigner $3j$-symbol.[22] This formula is valid if the atom–molecule collision time $\tau_{\rm col}$ is less than the molecule's rotation period $\tau_{\rm rot}$ (see Ref. 23),

$$\frac{\tau_{\rm col}}{\tau_{\rm rot}}\ll 1.\tag{23}$$

For diatomic molecules this condition becomes[24,25]

$$\sqrt{\frac{\mu\Delta\varepsilon}{M_r kT}}\ll 1,\tag{24}$$

where μ is the reduced mass of the collision partners, M_r is the reduced mass of the atoms comprising the diatomic molecule, and $\Delta\varepsilon=|\Delta\varepsilon_\alpha(J_{\alpha 1}J_\alpha)|$.

Khare[20] derived Eq. (22) in the (energy) sudden approximation, i.e., by replacing the rotational energy operator of the molecule with a constant. This energy approximation ignores the dependence of the cross section on the energy in the outgoing channel. The factor $1+\varepsilon_\alpha(J_\alpha)/E$ in (22) appears by virtue of detailed balance.[20,24] The sudden approximation is effective when the change in rotational energy is small compared to the total energy.

Taking into account the orthogonality relation[20]

$$\sum_{J_{\alpha 1}}(2J_{\alpha 1}+1)\begin{pmatrix}J_\alpha & J_{\alpha 1} & L\\ 0 & 0 & 0\end{pmatrix}^2=1\tag{25}$$

and combining (21) with (22), we obtain

$$\sigma^t_\alpha(E,J_\alpha)=\left[1+\frac{\varepsilon_\alpha(J_\alpha)}{E}\right]\sigma^t_{0\alpha}(E+\varepsilon_\alpha(J_\alpha)),\tag{26}$$

where

$$\sigma^t_{0\alpha}(E+\varepsilon_\alpha(J_\alpha))=\int(1-\cos\theta)\left(\sum_L\sigma_\alpha\times(E+\varepsilon_\alpha(J_\alpha),\theta,\phi;\ 0\to L)\right)d\Omega.\tag{27}$$

Since in the sudden approximation it is assumed that there is only a small change in rotational energy, Eq. (26) is valid if $\varepsilon_\alpha(J_\alpha)\ll E$. In an approximation that is linear in the small parameter

$$\frac{\varepsilon_\alpha(J_\alpha)}{kT}\ll 1,\tag{28}$$

Eqs. (19) and (26) yield

$$\nu^t_\alpha(J_\alpha)=\nu^{\rm vib}_{0\alpha}+\frac{\varepsilon_\alpha(J_\alpha)}{kT}\nu^{\rm vib}_{1\alpha},\tag{29}$$

where

$$\nu^{\rm vib}_{0\alpha}=\frac{2a}{3(kT)^3}\int_0^\infty E^2\exp\left(-\frac{E}{kT}\right)\sigma^t_{0\alpha}(E)dE,\tag{30}$$

$$\nu^{\rm vib}_{1\alpha}=\nu^{\rm vib}_{0\alpha}-\frac{2a}{3(kT)^2}\int_0^\infty E\exp\left(-\frac{E}{kT}\right)\sigma^t_{0\alpha}(E)dE,\tag{31}$$

$$\sigma^t_{0\alpha}(E)=\int(1-\cos\theta)\left(\sum_L\sigma_\alpha(E,\theta,\phi;\ 0\to L)\right)d\Omega.\tag{32}$$

The quantities $\nu^{\rm vib}_{0\alpha}$ and $\nu^{\rm vib}_{1\alpha}$ in (29) have the dimensions of collision rate and depend only on the vibrational state α. All of the dependence of $\nu^t_\alpha(J_\alpha)$ on the vibrational state J_α is in the factor $\varepsilon_\alpha(J_\alpha)/kT$. The quantity $\sigma^t_{0\alpha}(E)$ in (32) is the total (i.e., elastic and inelastic) transport scattering cross section of a molecule in vibrational level α and rotational state $J_\alpha=0$.

We now simplify Eq. (20) for $\nu^c_\alpha(J_\alpha)$. Since the main contribution to the integral in (20) is provided by energies $E\sim kT$, in view of condition (28) the square root in the integrand can be expanded in a power series in the small quantity $\Delta\varepsilon_\alpha(J_{\alpha 1}J_\alpha)/E$ (energies $E\to 0$ contribute essentially nothing to the integral). In an approximation linear in the small parameter

$$\frac{|\Delta\varepsilon_\alpha(J_{\alpha 1}J_\alpha)|}{kT}\ll 1,\tag{33}$$

Eq. (20) yields

$$\nu^c_\alpha(J_\alpha)=\frac{a}{3(kT)^3}\sum_{J_{\alpha 1}}\Delta\varepsilon_\alpha(J_{\alpha 1}J_\alpha)\times\int_0^\infty E\exp\left(-\frac{E}{kT}\right)\sigma^c_\alpha(E;J_\alpha\to J_{\alpha 1})\,dE.\tag{34}$$

Allowing for the correction term $\varepsilon_\alpha(J_\alpha)$ in the linear approximation in (22), substituting (22) into (34) would exceed the numberical precision. Hence we must plug (22) into (34) with $\varepsilon_\alpha(J_\alpha)=0$. We also allow for the fact that for linear molecules the energy of the rotational level J_α is

$$\varepsilon_\alpha(J_\alpha)=B_\alpha J_\alpha(J_\alpha+1),\tag{35}$$

where B_α is the rotational constant for vibrational level α. Next, using the fact that[26,27]

$$\sum_{J_{\alpha 1}}J_{\alpha 1}(J_{\alpha 1}+1)(2J_{\alpha 1}+1)\begin{pmatrix}J_\alpha & J_{\alpha 1} & L\\ 0 & 0 & 0\end{pmatrix}^2=J_\alpha(J_\alpha+1)+L(L+1)\tag{36}$$

and taking (25) into account, we obtain from (34)

$$\nu_\alpha^c(J_\alpha)=\nu_\alpha^c\equiv\frac{a}{3(kT)^3}\sum_L\varepsilon_\alpha(L)\times\int_0^\infty E\,\exp\left(-\frac{E}{kT}\right)\sigma_\alpha^c(E;0\to L)dE. \tag{37}$$

It is clear from (37) that in the linear approximation, $\nu_\alpha^c(J_\alpha)$ is independent of the initial rotational state of the molecule.

Thus, if for the overwhelming majority of molecules the conditions (24), (28), and (33) are met, then to a first approximation the transport collision rate $\nu_\alpha(J_\alpha)$ in (17) is

$$\nu_\alpha(J_\alpha)=\nu_\alpha^{\rm vib}+\frac{\varepsilon_\alpha(J_\alpha)}{kT}\nu_{1\alpha}^{\rm vib}, \tag{38}$$

where $\nu_\alpha^{\rm vib}\equiv\nu_{0\alpha}^{\rm vib}+\nu_\alpha^c$ and $\nu_\alpha^{\rm vib}$ depend only on the vibrational state α. The dependence of the transport rate $\nu_\alpha(J_\alpha)$ on the initial rotational state J_α is due only to the factor $\varepsilon_\alpha(J_\alpha)/kT$.

4. RELATIVE DIFFERENCE OF TRANSPORT COLLISION RATES

Suppose that the light is in resonance with the rovibrational transition $nJ_i\to mJ_f$. Then for the relative difference of the collision rates that enter into Eq. (6) for the LID velocity, Eqs. (16) and (38) yield

$$\frac{\Delta\nu}{\nu}\equiv\frac{\nu_m-\nu_n}{\nu_n}=\frac{\nu_m(J_f)-\nu_n(J_i)}{\nu_n(J_i)}\approx\frac{\nu_m(J_f)-\nu_n(J_i)}{\nu_n^{\rm vib}}$$
$$=\frac{\nu_m^{\rm vib}-\nu_n^{\rm vib}}{\nu_n^{\rm vib}}+\frac{\varepsilon_m(J_f)\nu_{1m}^{\rm vib}-\varepsilon_n(J_i)\nu_{1n}^{\rm vib}}{kT\nu_n^{\rm vib}}. \tag{39}$$

Here we can neglect the difference between the values of the rotational energy $\varepsilon_m(J)$ and $\varepsilon_n(J)$ with the same J in different vibrational states, since it does not exceed several percent,[28] which is less than the numerical precision. Since experiments have shown that the relative difference of the transport collision rates of molecules in the vibrational ground and excited states, $|\nu_m^{\rm vib}-\nu_n^{\rm vib}|/\nu_n^{\rm vib}$, is usually $\lesssim 1\%$ (see Refs. 7–9), in (39) we also neglect the difference between $\nu_{1m}^{\rm vib}$ and $\nu_{1n}^{\rm vib}$ because it is expected that $|\nu_{1m}^{\rm vib}-\nu_{1n}^{\rm vib}|/\nu_n^{\rm vib}$ is also $\lesssim 1\%$. As a result, from (39) we obtain

$$\frac{\Delta\nu}{\nu}=\left(\frac{\Delta\nu}{\nu}\right)_{\rm vib}+\left(\frac{\Delta\nu}{\nu}\right)_{\rm rot}, \tag{40}$$

where

$$\left(\frac{\Delta\nu}{\nu}\right)_{\rm vib}=\frac{\nu_m^{\rm vib}-\nu_n^{\rm vib}}{\nu_n^{\rm vib}}, \tag{41}$$
$$\left(\frac{\Delta\nu}{\nu}\right)_{\rm rot}=\frac{\varepsilon(J_f)-\varepsilon(J_i)}{kT}\frac{\nu_{1n}^{\rm vib}}{\nu_n^{\rm vib}}.$$

Here $\varepsilon(J)=BJ(J+1)$, with B the rotational constant. Thus, the factor $\Delta\nu/\nu$ is a sum of two independent terms, the vibrational term $(\Delta\nu/\nu)_{\rm vib}$ and the rotational term $(\Delta\nu/\nu)_{\rm rot}$. The vibrational term $(\Delta\nu/\nu)_{\rm vib}$ depends only on the vibrational numbers m and n, while the rotational term $(\Delta\nu/\nu)_{\rm rot}$ depends only on the rotational numbers J_i and J_f.

The representation (40) of the factor $\Delta\nu/\nu$ as a sum of independent vibrational and rotational terms was suggested on qualitative grounds by Chapovsky *et al.*[7] and has been used to process the data of LID experiments.[7–9,29]

The quantity $\delta\equiv\nu_{1n}^{\rm vib}/\nu_n^{\rm vib}$ in (41) can be calculated if we assume that the dependence of the total transport cross section $\sigma_{0n}^t(E)$ on the kinetic energy is given by a power law: $\sigma_{0n}^t(E)\propto E^{-2/n}$. In elastic scattering, such an energy dependence of the cross section corresponds to a power-law interaction potential $U\propto r^{-n}$ (see Ref. 30). Since when (33) holds, inelastic rotational transitions have a small effect on the path of the colliding particles, for the total transport cross section the dependence $\sigma_{0n}^t(E)\propto ED^{-2/n}$ also approximately corresponds to the potential $U\propto r^{-n}$.

Substituting $\sigma_{0n}^t(E)\propto E^{-2/n}$ into (30) and (31) yields

$$\delta\approx\frac{\nu_{1n}^{\rm vib}}{\nu_{0n}^{\rm vib}}=\frac{(1-2/n)\Gamma(2-2/n)}{\Gamma(3-2/n)}, \tag{42}$$

where $\Gamma(x)$ is the gamma function. Since the value $n\gg 1$ is usually most suitable for describing realistic interaction potentials, Eq. (42) yields $\delta\approx 0.5$. Thus, for sufficiently short-range interaction potentials,

$$\left(\frac{\Delta\nu}{\nu}\right)_{\rm rot}\approx\frac{\varepsilon(J_f)-\varepsilon(J_i)}{2kT}. \tag{43}$$

In the transitions $P(J_i)$ (i.e., with $J_i\to J_f=J_i-1$) and $R(J_i-1)$ (i.e., with $J_i-1\to J_f=J_i$), the absolute value of the factor $(\Delta\nu/\nu)_{\rm rot}$ is the same but the signs are different:

$$\left(\frac{\Delta\nu}{\nu}\right)_{{\rm rot}\,R(J_i-1)}=-\left(\frac{\Delta\nu}{\nu}\right)_{{\rm rot}\,P(J_i)}=\frac{2BJ_i}{kT}\frac{\nu_{1n}^{\rm vib}}{\nu_n^{\rm vib}}\approx\frac{BJ_i}{kT}. \tag{44}$$

The sum of the relative differences of the collision rates $\Delta\nu/\nu$ for the transitions $P(J_i)$ and $R(J_i-1)$ does not contain the rotational term $(\Delta\nu/\nu)_{\rm rot}$:

$$\left(\frac{\Delta\nu}{\nu}\right)_{P(J_i)}+\left(\frac{\Delta\nu}{\nu}\right)_{R(J_i-1)}=2\left(\frac{\Delta\nu}{\nu}\right)_{\rm vib}. \tag{45}$$

On the other hand, the difference of the relative collision rates does not contain the vibrational term $(\Delta\nu/\nu)_{\rm vib}$:

$$\left(\frac{\Delta\nu}{\nu}\right)_{P(J_i)}-\left(\frac{\Delta\nu}{\nu}\right)_{R(J_i-1)}=2\left(\frac{\Delta\nu}{\nu}\right)_{{\rm rot}\,P(J_i)}. \tag{46}$$

To give a numerical example, at $B=0.5\ {\rm cm}^{-1}$ and $T=300\ {\rm K}$ we have $(\Delta\nu/\nu)_{{\rm rot}\,P(1)}\approx -0.25\%$ and $(\Delta\nu/\nu)_{{\rm rot}\,P(4)}\approx -1\%$. For a typical value $(\Delta\nu/\nu)_{\rm vib}\approx 1\%$ (Refs. 7–9), we then have $\Delta\nu/\nu\approx 0$ for the $P(4)$ transition and, according to Ref. 14, anomalous LID is to be expected.[7,15–17]

5. CONCLUSION

We have derived, in the (energy) sudden approximation, a simple analytic dependence (Eqs. (40) and (41)) for the relative difference of transport collision rates $\Delta\nu/\nu$ of collisions of linear molecules and atoms on the rotational numbers of the combining levels of the molecules (i.e., those

affected by radiation). The factor $\Delta\nu/\nu$ measured in molecular LID experiments was found to be equal to the sum of independent vibrational $(\Delta\nu/\nu)_{\rm vib}$ and rotational $(\Delta\nu/\nu)_{\rm rot}$ terms. For the rotational term, the distinguishability of the buffer particles is unimportant in view of the weak dependence of $\nu_{1n}^{\rm vib}/\nu_n^{\rm vib}$ in (41) on the details of the interaction potential. It is to be expected that at room temperature the derived expressions are valid for linear molecules with moderate values of the rotational constant $B\lesssim 1\ {\rm cm}^{-1}$ and initial rotational number J_i, since the sudden approximation presupposes a small variation in rotational energy in comparison to the kinetic energy of the relative motion of the colliding particles.

In LID experiments involving molecules with a small value of the rotational constant, it is possible to verify Eq. (41) for $(\Delta\nu/\nu)_{\rm rot}$ based on the factorization formula (22) for cross sections. LID experiments are significantly less complicated than experiments with molecular beams, and establishing the accuracy and limits of applicability of the factorization formula (22) for cross sections in LID experiments is of some interest for the physics of intermolecular interactions.

Among linear molecules, only HF has been studied in LID experiments.[8,9] For this molecule the condition (24) of applicability of the sudden approximation is not met because of the large rotational constant, $B\approx 21\ {\rm cm}^{-1}$. Nevertheless, for moderate values $J_i=1,2,3$ at which the rotational energy $\varepsilon(J_i)\lesssim kT$, Eq. (44) describes the trends in the behavior of $(\Delta\nu/\nu)_{\rm rot}$ with increasing J_i fairly well for the transitions $P(J_i)$ and $R(J_i-1)$ (in the experiments described in Refs. 8 and 9, for the mixtures HF–Ar, HF–Kr, and HF–Xe an increase in J_i for $J_i=1,2,3$ was found to lead to an almost linear decrease in $(\Delta\nu/\nu)_{{\rm rot}\,P(J_i)}$ and an almost linear increase in $(\Delta\nu/\nu)_{{\rm rot}\,R(J_i-1)}$, in accordance with (44)).

For the HF molecule, Eq. (44) yields a rate of variation of $(\Delta\nu/\nu)_{\rm rot}$ as a function of the initial rotational number J_i that is several times greater than the actual value. The reason is that collision rotational transitions, which are responsible for the dependence of $(\Delta\nu/\nu)_{\rm rot}$ on J_i, emerge solely because of the anisotropy (nonsphericity) of the intermolecular interaction potential. When condition (23) is violated, the angle through which the molecule rotates in the course of the collision (time $\tau_{\rm col}$) is not small, and the interaction potential is effectively averaged over angle. As a result, the effective anisotropy of the interaction potential decreases, which ensures a dependence of $(\Delta\nu/\nu)_{\rm rot}$ on J_i in HF that is weaker than the one predicted by (44).

In their LID experiment involving CH_3F molecules, Bloemink *et al.*,[29] also thoroughly studied of the dependence of $\Delta\nu/\nu$ on the rotational number. In Ref. 29, as with the HF molecule, the applicability condition (24) of the sudden approximation is not met. Although our results cannot be directly applied to symmetric top molecules (such as CH_3F), we can compare the experimental results with those obtained from Eq. (41) if the energy factor $\varepsilon(J_f)-\varepsilon(J_i)$ is replaced by $\varepsilon(J_f,K_f)-\varepsilon(J_i,K_i)$, where K is the projection of the angular momentum on the axis of the top. With this change, Eq. (41) correctly describes trends in the behavior of $\Delta\nu/\nu$ as a function of the rotational number K_i (an almost linear dependence of $\Delta\nu/\nu$ on K_i was observed in the experiment), and for $J_i\geqslant 11$ Eq. (41) also correctly describes the behavior of $\Delta\nu/\nu$ as a function of the rotational number J_i.

The present work was sponsored by the Russian Fund for Fundamental Research (Grant No. 96-02-19556) and the State Scientific and Technical Program (Laser Physics, Grant No. 7.41).

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

[1] A. M. Shalagin, ''Light-induced drift,'' in *Physics Encyclopedia* [in Russian], Vol. 4, A. M. Prokhorov (ed.), Bol'shaya Rosiĭskaya Éntsiklopediya, Moscow (1994), p. 468.
[2] F. Kh. Gel'mukhanov and A. M. Shalagin, JETP Lett. **29**, 711 (1979).
[3] G. Nienhuis, Phys. Rep. **138**, 151 (1986).
[4] H. G. C. Werij and J. P. Woerdman, Phys. Rep. **169**, 145 (1988).
[5] S. G. Rautian and A. M. Shalagin, *Kinetic Problems of Nonlinear Spectroscopy*, North-Holland, Amsterdam (1991).
[6] E. R. Eliel, Adv. At., Mol., Opt. Phys. **30**, 199 (1992).
[7] P. L. Chapovsky, G. J. van der Meer, J. Smeets, and L. J. F. Hermans, Phys. Rev. A **45**, 8011 (1992).
[8] E. J. van Duijn, H. I. Bloemink, E. R. Eliel, and L. J. F. Hermans, Phys. Lett. A **184**, 93 (1993).
[9] E. J. van Duijn, R. Nokhai, and L. J. F. Hermans, J. Chem. Phys. **105**, 6375 (1996).
[10] F. Kh. Gel'mukhanov and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **78**, 1674 (1980) [Sov. Phys. JETP **51**, 839 (1980)].
[11] V. P. Mironenko and A. M. Shalagin, Izv. Akad. Nauk SSSR, Ser. Fiz. **45**, 995 (1981).
[12] S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in the Spectra of Atoms and Molecules* [in Russian], Nauka, Novosibirsk (1979).
[13] P. L. Chapovskiĭ, Izv. Akad. Nauk SSSR, Ser. Fiz. **53**, 1069 (1989).
[14] R. W. M. Hoogeveen, G. J. van der Meer, L. J. F. Hermans, and P. L. Chapovsky, J. Chem. Phys. **90**, 6143 (1989).
[15] L. J. F. Hermans, Int. Rev. Phys. Chem. **11**, 289 (1992).
[16] G. J. van der Meer, J. Smeets, S. P. Pod'yachev, and L. J. F. Hermans, Phys. Rev. A **45**, R1303 (1992).
[17] F. Kh. Gel'mukhanov and A. O. Parkhomenko, Phys. Lett. A **162**, 45 (1992).
[18] R. Goldflam, D. J. Kouri, and S. Green, J. Chem. Phys. **67**, 5661 (1977).
[19] G. A. Parker and R. T. Pack, J. Chem. Phys. **67**, 1585 (1977).
[20] V. Khare, J. Chem. Phys. **68**, 4631 (1978).
[21] M. Faubel, Adv. At. Mol. Phys. **30**, 199 (1992).
[22] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum*, World Scientific, Singapore (1987).
[23] A. S. Dickinson, Comput. Phys. Commun. **17**, 51 (1979).
[24] Z. H. Top and D. J. Kouri, Chem. Phys. **37**, 265 (1979).
[25] A. V. Bogdanov, G. V. Dubrovskiĭ, and A. I. Osipov, Khim. Fiz. **4**, 1155 (1985).
[26] M. L. Strekalov, Khim. Fiz. **7**, 1182 (1988).
[27] A. V. Storozhev and M. L. Strekalov, Khim. Fiz. **16**, No. 2, 17 (1997).
[28] A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions*, Springer, Berlin (1985).
[29] H. I. Bloemink, J. M. Boon-Engering, P. L. Chapovsky, L. J. F. Hermans, and E. R. Eliel, J. Phys. B **27**, 4559 (1994).
[30] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed., Pergamon Press, Oxford (1976).

Translated by Eugene Yankovsky

Nonexponential temperature dependence of the rate of threshold inelastic processes in dense media

N. L. Aleksandrov*) and A. N. Starostin

Moscow Physicotechnical Institute, 141700 Dolgoprudnyi, Moscow Region, Russia
(Submitted 17 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1661–1674 (May 1998)

The influence of the density of the medium on the temperature dependence of the rate constants of inelastic processes is investigated. It is shown that besides the effects like lowering of the ionization potential, which accelerate excitation and ionization processes in a nonideal plasma, there is a stronger mechanism for such acceleration, which is associated with the high frequency of collisions between particles and leads to destruction of the one-to-one relation between the energy and momentum of the particles in a dense medium. It is manifested by the presence of power-law tails in the equilibrium momentum distribution of the particles, which leads to a nonexponential temperature dependence of the rates of inelastic reactions in dense gases and nonideal plasmas. A kinetic equation for the generalized energy and momentum distribution function of electrons in an external electric field, which permits investigation of the effect under consideration under nonequilibrium conditions, is presented.
 [S1063-7761(98)00805-1]

1. INTRODUCTION

The temporal relaxation of the density of particles and excited states in different media is described by the corresponding balance equations. The coefficients in these equations, which may be called rate constants, usually do not depend on the particle density. However, in dense nonideal media the strong interparticle interaction causes the rates of elementary processes to depend on the density of the medium. This phenomenon is most pronounced for threshold processes. Its study (mainly by theoretical methods) has been confined to the electron-impact ionization and excitation of atoms (and the reverse reactions) and electron attachment to molecules.[1–10] However, the nonideal behavior caused by both the Coulomb interaction in a plasma[1–6] and by electron–atom and atom–atom interactions as electrons move in dense gases and liquids[7–10] has also been considered.

It is presently believed that the main reason for the variation of the rate of threshold electron processes as the density of the medium increases is the lowering of their energy threshold. The temperature dependence of the rate constant remains exponential, although it can fail to coincide with the Arrhenius law because of a possible temperature dependence of the decrease in the ionization potential. Other effects, such as degeneracy of the electron gas in a plasma[4] or the appearance of a structure for the medium in a liquid[7,8] are assumed to be less important. The systematic theoretical approach to the determination of the lowering of the energy threshold of elementary processes is based on the formalism of kinetic Green's functions.[11] Under such a treatment the real part of the density shift of the energy is taken into account, and the imaginary part is neglected, i.e., it is assumed that there is a one-to-one relationship between the energy and momentum of the particles. At the same time, it is known that in dense media the quantum uncertainty of the energy of particles caused by their frequent collisions destroys the one-to-one relationship between the energy and momentum of the particles.[12] In this case a generalized energy and momentum distribution function of the particles and scattering cross sections of the particles outside the mass surface must be introduced. This formalism has previously been used to describe resonant radiation transfer in dense absorbing media.[13] It has been shown in reference to particles that the quantum uncertainty of the energy arising as a result of collisions leads to the appearance of power-law tails in the particle momentum distribution even under thermodynamic equilibrium conditions.[14]

The purpose of the present work is to theoretically study the temperature dependence of the rates of threshold processes in dense media with consideration of the effects resulting from the quantum uncertainty of the energy of particles caused by interparticle interactions.

2. QUALITATIVE TREATMENT

As was shown in Ref. 13, the radiative energy transfer in dense, strongly absorbing media, in which the mean free path of the resonant ''photon'' is commensurate with its wavelength, should be described by the generalized spectral intensity $J(\omega,\mathbf{k})$, in which the frequency ω and the wave vector $\mathbf{k}$ are independent variables. In Ref. 13 equations which are satisfied by $J(\omega,\mathbf{k})$ were presented, and some consequences of the solution of these equations, some of which could be proved experimentally,[15,16] were considered. The machinery of the generalized energy and momentum distribution functions $f(E,\mathbf{p})$ as applied to particles was considered in Refs. 12, 14, 17, and 18. Let us recall the basic definitions introduced in those studies. For example, for $f(E,\mathbf{p})$ we have

$$f(E,\mathbf{p})=\frac{1}{2\pi}\int d\tau\, d^3\rho\, \exp(iE\tau-i\mathbf{p}\cdot\boldsymbol{\rho})$$
$$\times\langle\hat{\Psi}^+(x_2)\hat{\Psi}(x_1)\rangle. \quad (1)$$

Here $\tau=t_1-t_2$, $\boldsymbol{\rho}=\mathbf{r}_1-\mathbf{r}_2$, and $\hat{\Psi}(x)$ is the field operator of a particle in the Heisenberg representation, and the angle brackets denote quantum-statistical averaging. To simplify the writing of the equations we shall assume, for now, that $\hbar=1$. By definition, the energy (frequency) E and the momentum $\mathbf{p}$ in (1) are the independent variables for Fourier transformation with respect to the time and coordinate differences, and, for example, the values of E over the entire real axis from $-\infty$ to $+\infty$ should be considered. The function $f(E,\mathbf{p})$ introduced in this manner is a generalization of the Wigner distribution (see Refs. 12 and 18, where the physical meaning of this distribution in the quasiclassical limit is explained). Integrating (1) over the energy or over the momentum, we can obtain the distribution with respect to the momentum alone, $f(\mathbf{p})$, or the distribution with respect to the energy alone, $f(E)$. Under the conditions of thermodynamic equilibrium we can obtain[12,14,18]

$$f(E,\mathbf{p})=-i\frac{n(E)}{2\pi}(G^R(E,\mathbf{p})-G^A(E,\mathbf{p})). \quad (2)$$

In (2) $n(E)$ denotes the equilibrium occupation numbers for the particles. For example, for fermions

$$n(E)=\frac{1}{e^{(E-\mu)/T}+1}. \quad (3)$$

In addition, G^R and G^A are the retarded and advanced Green's functions, which are related by

$$G^A=(G^R)^*.$$

For $G^R(E,\mathbf{p})$ we can write the expression

$$G^R(E,\mathbf{p})=\frac{1}{E-\varepsilon_{\mathbf{p}}-\Sigma^R(E,\mathbf{p})}, \quad (4)$$

where $\varepsilon_{\mathbf{p}}=\mathbf{p}^2/2M$, M is the mass of a particle, and the mass operator Σ^R is consistent with the rules of the diagram formalism.[12,14,17,18] The quantity $\mathrm{Re}\,\Sigma^R$ specifies the shift of the kinetic energy of a particle due to its interaction with the surrounding neighbors, and $\mathrm{Im}\,\Sigma^R$ specifies the width or damping of an excitation with the energy E and the momentum $\mathbf{p}$. For example, in a weakly nonideal plasma the electron energy shift (in the present case it is the renormalization of the electron chemical potential μ) can be represented in the form[11]

$$\mathrm{Re}\,\Sigma^R=-\frac{\Gamma}{2}T, \quad (5)$$

where $\Gamma=e^2/r_DT$ is the interaction parameter, r_D is the Debye radius, T is the temperature, and e is the charge of an electron. It was shown in Refs. 1–8 during an analysis of the influence of density effects on the ionization rate constant that including $\mathrm{Re}\,\Sigma^R$ lowers the ionization potential in dense media (the ionization potential lowering mechanism); in a nonideal plasma this causes exponential acceleration of the reaction ($\propto e^{\Gamma}$). It should be noted here that the expression (5) can be rigorously obtained only in the limit $\Gamma\lesssim 1$ and that, from the standpoint of this theory, the quantitatively appreciable acceleration of the reaction observed for $\Gamma>1$ results from extrapolating its results beyond its range of applicability. The width $\mathrm{Im}\,\Sigma^R$ was neglected in the studies just cited.[1–8] In the gas approximation we can obtain the estimate

$$\mathrm{Im}\,\Sigma^R\propto\nu\propto N\sigma v,$$

where ν is the collision frequency, N is the particle density, σ is the scattering cross section, and v is the velocity. For Coulomb scattering we have

$$\mathrm{Im}\,\Sigma^R\propto\frac{\lambda\!\!\!^{-}}{r_D}\Gamma T. \quad (6)$$

Here $\lambda\!\!\!^{-}$ is the thermal de Broglie wavelength of the particle, i.e., $\mathrm{Im}\,\Sigma^R$ is determined by quantum effects. In a nonideal plasma the dimensionless parameters $\lambda\!\!\!^{-}/r_D$ and Γ are independent, and conditions under which $\mathrm{Im}\,\Sigma^R\propto\mathrm{Re}\,\Sigma^R$ holds are easily created.

In the low-density limit, where we have $\nu/T\ll 1$ and $\mathrm{Re}\,\Sigma^R/T\ll 1$, from (2) we can obtain

$$f(E,\mathbf{p})=n(E)\,\delta\left(E-\frac{\mathbf{p}^2}{2M}\right). \quad (7)$$

In this case the energy and momentum are uniquely related by the dispersion relation for a free particle $E=\mathbf{p}^2/2M$, and, just as for photons in a rarefied gas, the following relation holds:[13,18]

$$J(\omega,\mathbf{k})=n(\omega)\,\delta(\omega^2-c^2k^2). \quad (8)$$

The interaction effects broaden the δ functions in (7) and (8) and shift the particle or photon energies. Thus, instead of (7), from (2) and (4) we obtain an analog of the Lorentzian line shape:

$$f(E,\mathbf{p})=\frac{n(E)}{\pi}\frac{\mathrm{Im}\,\Sigma^R}{(E-\varepsilon_{\mathbf{p}}-\mathrm{Re}\,\Sigma^R)^2+(\mathrm{Im}\,\Sigma^R)^2}. \quad (9)$$

Because of the finite value of the width $\mathrm{Im}\,\Sigma^R$, it follows from (9) that for a fixed value of the energy (frequency) E the magnitude of the momentum can take any value, i.e., in a dense medium the dispersion of a free particle $E=p^2/2M$ spreads out because of the frequent collisions, and the momentum distribution contains power-law tails, whose existence was first discussed in Ref. 14. Thus, using the gas approximation to calculate Σ^R[14] and integrating (9) over the energy, for nondegenerate particles we obtain

$$f(\mathbf{p})=\int_{-\infty}^{\infty}dE\; f(E,\mathbf{p})=f_M(\mathbf{p})+\frac{\hbar\,\nu_{\mathbf{p}}T}{2\pi\varepsilon_{\mathbf{p}}^2}e^{\mu/T}, \quad (10)$$

where $f_M(\mathbf{p})$ is the Maxwellian momentum distribution of the particles and $\nu_{\mathbf{p}}$ is the collision frequency, which depends on the momentum $\mathbf{p}$. The quantum nature of the effect is underscored here by the explicit introduction of Planck's constant $\hbar$. In Ref. 14 an expression similar to (10) was obtained for degenerate particles, and the non-Maxwellian term was proportional to the square of the gas density. It should be noted that the presence of power-law tails in the

equilibrium particle momentum distribution can be obtained explicitly only in the gas approximation, i.e., when the approximation of binary collisions is used for $\mathrm{Im}\,\Sigma^R(E,\mathbf{p})$ (see Ref. 14). In the general case of nonideal media of high-density gases, where we have $\hbar\nu/T \geqslant 1$, this approximation can be violated, and finding quantitative results becomes practically impossible.

However, in the limit of a relatively low density the presence of the power-law tails can have a significant influence on the rate constants of processes with a large energy threshold. To underline the importance of this effect, we note that the particle energy distribution obtained from (9) by integrating over the momentum remains exponential, even on distant tails. For example, when $\nu=\nu_{\mathbf{p}}=\mathrm{const}$ holds, for $f(E)$ at $E>0$ we obtain

$$f(E)=e^{(\mu-E)/T}\,\mathrm{Re}\sqrt{E+i\nu}. \tag{11}$$

To ascertain the rate constants of the processes, the generalized cross sections outside the mass surface [which depend both on the momenta before and after scattering and on the energy E, which is not related to the momentum by expressions of the form (7)] should be integrated in the general case with the generalized distribution functions (1) defined above (for further details, see the next section). For example, in the Born approximation for the case of a static particle interaction potential, the scattering amplitude and cross section are specified by the Fourier component of the potential and depend only on the difference between the momenta. When the scattering cross section is averaged, it is obvious in this case that the result is determined by a momentum distribution function of the type (10) and can differ strongly, by an order of magnitude, from the standard result, which takes into account only the Maxwellian distribution, i.e., the first term in (10), for large momenta.

The effects associated with the presence of power-law tails in the particle momentum distribution can lead to exponential acceleration of reactions, such as the electron-impact excitation and ionization of atoms and ions, electron attachment to molecules, the $V-T$ relaxation of molecules, and chemical and nuclear conversions in dense gases and nonideal plasmas. Moreover, these effects are not associated directly with the ionization potential lowering mechanism, which was essentially taken into account only in Refs. 1–8, and are manifested for $\Gamma \leqslant 1$, where the theory is still applicable.

3. RATE CONSTANTS OF INELASTIC PROCESSES IN DENSE MEDIA

Let us consider the general expression for the rate constant of a process in which the colliding particles e and a, which are initially in states $i(e)$ and $\alpha(a)$, pass into states j and β. In the Born approximation this process is described by the diagram $\Sigma^{+-}_{i\alpha,j\beta}G^{+-}_{\alpha}$,[17,18] which is presented in Fig. 1. For the averaged quantity $n_e n_a k_{i\alpha,j\beta}$ we can use the equilibrium values of the kinetic Green's functions[12,17,18] and this diagram to obtain

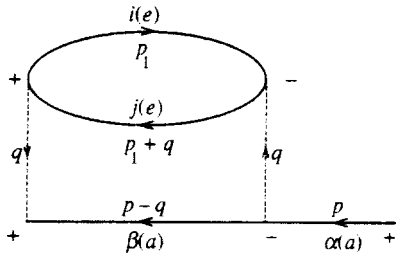

FIG. 1. Diagram of a process in the Born approximation.

$$\begin{aligned}\overline{n_e n_a k_{i\alpha,j\beta}} = \int \frac{d^4p\,d^4p_1\,d^4q}{(2\pi)^{12}}|V_{i\alpha,j\beta}(q)|^2 n_e(\omega_1)\\ \times\delta_\gamma(\omega_1-\varepsilon_{\mathbf{p}_1})(1-n_e(\omega_1+\omega))\\ \times\delta_\gamma(\omega_1+\omega-\omega_{\beta\alpha}-\varepsilon_{\mathbf{p}_1+\mathbf{q}})N_a(\omega_{\mathbf{p}})\\ \times\delta_\gamma(\omega_p-E_{\mathbf{p}})(1-N_a(\omega_p-\omega))\\ \times\delta_\gamma(\omega_p-\omega+\omega_{ij}-E_{\mathbf{p}-\mathbf{q}}).\end{aligned} \tag{12}$$

In (12) we introduced the notation

$$n_e(\omega)=\frac{1}{e^{(\omega-\mu_e)/T}+1},\quad N_a(\omega)=\frac{1}{e^{(\omega-\mu_a)/T}+1},$$

$$\varepsilon_{\mathbf{p}}=\frac{\mathbf{p}^2}{2m_e},\quad E_{\mathbf{p}}=\frac{\mathbf{p}^2}{2M_a},$$

and $V_{i\alpha,j\beta}(q)$ is the Fourier transform of the particle interaction potential matrix element for the process $i(e),\alpha(a)\to j(e),\beta(a)$. In (12) we used the four-dimensional vectors $p(\mathbf{p},\omega_p)$, $p_1(\mathbf{p}_1,\omega_1)$, and $q(\mathbf{q},\omega)$. The quantity $\delta_\gamma(\omega-\varepsilon_{\mathbf{p}})$ has the form (compare Ref. 19)

$$\delta_\gamma(\omega-\varepsilon_{\mathbf{p}})=\frac{\gamma(\omega,\mathbf{p})}{\pi[(\omega-\varepsilon_{\mathbf{p}}-\Delta(\omega,\mathbf{p}))^2+\gamma^2(\omega,\mathbf{p})]}. \tag{13}$$

Formula (13) was written in analogy to (9), the width γ is equivalent to $\mathrm{Im}\,\Sigma^R$, and the shift $\Delta=\mathrm{Re}\,\Sigma^R$. In the limit $\gamma,\Delta\to 0$ Eq. (12) transforms into the ordinary expression for the rate constant, if the Pauli effects are omitted, i.e., if we assume that the occupation numbers are small and neglect them in comparison to unity:

$$\begin{aligned}\overline{n_e n_a k_{i\alpha,j\beta}} = \int \frac{d^3p\,d^3p_1\,d^3q}{(2\pi)^{12}}|V_{i\alpha,j\beta}(\mathbf{q})|^2 n_e(\varepsilon_{\mathbf{p}_1})\\ \times N_a(E_{\mathbf{p}})\,\delta(\varepsilon_{\mathbf{p}_1}+E_{\mathbf{p}}+\omega_{ij}\\ +\omega_{\alpha\beta}-\varepsilon_{\mathbf{p}_1+\mathbf{q}}-E_{\mathbf{p}-\mathbf{q}}).\end{aligned} \tag{14}$$

Thus, the general expression (12) contains the generalized distribution functions $f_e(\omega_1,\mathbf{p}_1)$ and $f_a(\omega_p,\mathbf{p})$ for colliding particles and takes into account the degeneracy and Pauli's principle (the multipliers of the form $1-n$). The ordinary δ functions expressing the energy conservation law have been replaced in (12) by Lorentzians of the form (13) with a finite width and a finite shift. In the general case the width γ is determined by all the interaction processes of a particular particle with the surrounding particles and must be found by

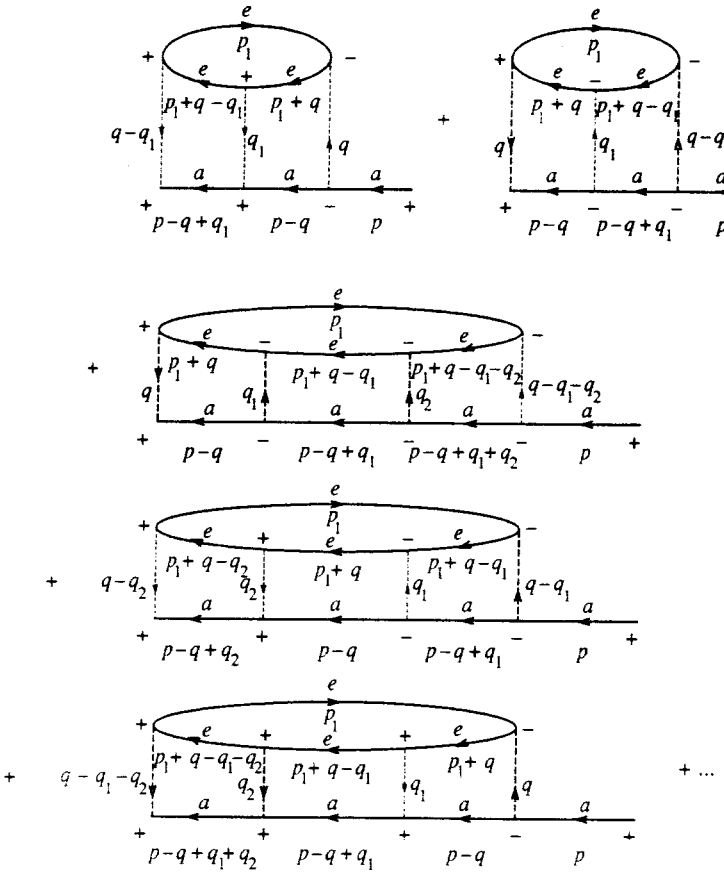

FIG. 2. Sum of ladder diagrams.

solving multidimensional nonlinear integral equations (compare Refs. 12, 19, and 20) consistently using expressions of the types (12) and (13), since

$$\gamma_e \equiv \mathrm{Im}\,\Sigma^R \propto \sum_{i\alpha,j\beta} \overline{n_a k_{i\alpha,j\beta}}.$$

If the colliding particles are not at equilibrium, a more general expression can be obtained from Fig. 1:

$$\overline{n_e n_a k_{i\alpha,j\beta}} = \int \Sigma^{+-}_{i\alpha,j\beta} G^{-+}_{a,\alpha} \frac{d^4p}{(2\pi)^4} = \int G^{-+}_{e,i}(p_1) G^{+-}_{a,\beta} \times (p-q) G^{+-}_{e,j}(p_1+q) G^{-+}_{a,\alpha}(p) |V_{i\alpha,j\beta}(q)|^2 \times \frac{d^4p\, d^4p_1\, d^4q}{(2\pi)^{12}}, \tag{15}$$

where the Green's functions $G^{\alpha\alpha'}(p)$ (Refs. 12, 17, and 18) for each kind of particle should be found from the solution of the generalized kinetic equations.[12,18–20] We note that the kinetic equations for the particle momentum distribution functions obtained by integrating the equations for the generalized distribution functions $f(E,\mathbf{p})$ [for the equivalent Green's functions $G^{-+}(E,\mathbf{p}) = 2\pi i f(E,\mathbf{p})$] over the energy are obtained in a closed form only in the approximation $\gamma \to 0$, i.e., when the $\delta_\gamma(\omega - \varepsilon_{\mathbf{p}})$ are replaced by ordinary δ functions.[12,17,18] A similar situation occurs in the theory of resonant radiation transfer in dense dispersive media, where a closed equation for J_ω cannot be obtained from the equation for the generalized spectral intensity $J(\omega,\mathbf{k})$ in the general case by integrating the equation for $J(\omega,\mathbf{k})$ over the wave vector.[13] An example of such a generalized kinetic equation for the electron distribution function $f(E,\mathbf{p})$ in a dense Lorentz gas is given in the next section.

The expression (15) can be generalized by dispensing with the Born approximation in describing a collision. For this purpose we should consider ladder diagrams with any number of rungs (see Fig. 2), but with one replacement of a minus sign by a plus sign in order to remain within the gas approximation of binary collisions.

It can be shown (compare Refs. 12, 21, and 22) that the square of the absolute value of the Fourier transform of the interaction potential matrix element V_q in (15) should be replaced by the square of the absolute value of the scattering amplitude of the particles in the medium $T(p,p_1,q)$, which satisfies the equation

$$T(p,p_1,q) = V_q + \int V_{q-q_1} G_e^{--}(p_1+q-q_1) G_a^{--} \times (p-q-q_1) T(p,p_1,q_1) \frac{d^4q_1}{(2\pi)^4}. \tag{16}$$

For simplicity in writing the equation, here we confine ourselves to elastic scattering. In the equilibrium case $G^{--}(p)$ can be written in the form[18]

$$G^{--}(p) = (1-n(E))G^R + n(E)G^A.$$

As was shown in Refs. 21 and 22, to find T in the gas approximation we can confine ourselves to the scattering amplitude of the particles in a vacuum, $T(p,p_1,q) \approx f(\mathbf{p},\mathbf{p}_1,\mathbf{p}-\mathbf{q},\mathbf{p}_1+\mathbf{q})$, which satisfies the ordinary Lippman–

Schwinger equation. In this approximation the scattering amplitude depends only on the relative movement of the particles in the center-of-mass system before and after scattering, where these momenta lie outside the mass surface, i.e.,

$$p_i \neq p_f, \quad \mathbf{p}_i = \frac{m_a \mathbf{p}_e - m_e \mathbf{p}_a}{m_a + m_e}, \quad \mathbf{p}_f = \mathbf{p}_i + \mathbf{q}.$$

The energy variable $E = \omega + \omega_1$ appears in the general expression for $T(p, p_1, q)$, but it vanishes in the approximation described, since the vacuum scattering amplitude outside the mass surface depends only on the three-dimensional relative momenta.[21,22]

To obtain qualitative and quantitative results, the expressions of the form (12) and (15) should be simplified. We assume that heavy nondegenerate particles of kind a interact weakly with other particles, i.e., $\gamma_a \to 0$, $\Delta_a \to 0$. For example, the particles of species a correspond to neutral atoms, whose density can be large, but $N_a \sigma_{aa} \nu_a \ll T$. The particles of species e correspond to electrons, for which we take into account the frequent collisions with neutrals and ions, so that $\gamma_e \neq 0$.

In this approximation it follows from (15) and (16) that

$$\overline{n_e k_{i,j}} = \int \frac{dE\ d^3p d^3p'}{(2\pi)^7} n(E)(1 - n(E-I)) |f_{i,j}(\mathbf{p}, \mathbf{p}')|^2 \times \delta_\gamma(E - \varepsilon_e(\mathbf{p}))\, \delta_\gamma(E - I - \varepsilon_e(\mathbf{p}')). \quad (17)$$

In (17) I corresponds to the threshold energy of the inelastic excitation or ionization process. The density shift of the energy can alter I, and its inclusion is thus equivalent to taking into account the ionization potential lowering mechanism.

We first consider the case of relatively low electron temperatures ($I/T \gg 1$) and use the approximation $f_{i,j}(\mathbf{p}, \mathbf{p}') \approx f_{ij} = \text{const}$ for a qualitative estimate. This approximation is fairly reasonable near the threshold of an inelastic process.[23] We neglect the degeneracy of the electrons and assume that the width is a constant, i.e., $\gamma(E, \mathbf{p}) = \gamma = \text{const}$. This provides an approximate description of the scattering of electrons on neutrals, for example, on He atoms, where $\sigma v = \text{const}$. In this model case the integration over $\mathbf{p}$ and $\mathbf{p}'$ in (17) can be performed explicitly, and for the excitation constant k_{ex} (the ionization constant k_{ion}) in the approximation just described we obtain

$$\overline{n_e k_{\text{ex(ion)}}} = A_{\text{ex(ion)}} \int dE\ n(E) \operatorname{Re}\sqrt{E + i\gamma}\ \operatorname{Re}\sqrt{E - I + i\gamma}. \quad (18)$$

Here $A_{\text{ex(ion)}}$ is a normalization constant, which also contains the value of the constant $|f_{ij}|^2$. In the limit $\gamma \to 0$ the ordinary near-threshold approximation for the rate constant of the process follows from (18):

$$\overline{n_e k_{\text{ex(ion)}}} = A_{\text{ex(ion)}} \int_I^\infty dE\ n(E) \sqrt{E} \sqrt{E - I}. \quad (19)$$

However, for $\gamma \neq 0$ it follows from (18) that values of E smaller than I make a nonzero contribution to the integral over the energy. In the present case this reflects the influence described above of the power-law tails in the particle momentum distribution. From (18) we can obtain the estimate

$$k_{\text{ex(ion)}} = k^{(0)}_{\text{ex(ion)}} \left(e^{-I/T} + \frac{\hbar \gamma}{2I} \right). \quad (20)$$

The second, nonexponential term in (20) can be small compared with unity (in the limit $\hbar\gamma \ll I$), but at low temperatures it can exceed the first term by many orders of magnitude. In this sense it can be stated that the reaction is exponentially accelerated in dense gases due to the presence of the power-law tails in the particle momentum distribution. As the temperature rises, the usual relation $k \propto e^{-I/T}$ also holds at low gas densities. Although consideration of the ionization potential lowering mechanism increases the reaction rate, it is less important under the present conditions than consideration of the finite width in (17). To estimate the analogous increase in the excitation (ionization) rate constant in a nonideal plasma, we numerically integrated (17) in the approximation $f(\mathbf{p}, \mathbf{p}') = f = \text{const}$, but with consideration of the dependence of the width γ on the momentum $\gamma(E, \mathbf{p}) \approx \gamma_{\mathbf{p}} = N_i \sigma_C v$, where N_i is the ion density and σ_C is the Coulomb transport scattering cross section [$\sigma_C \propto (\pi e^4/\varepsilon^2)\Lambda$, where Λ is the Coulomb logarithm]. Figures 3 and 4 present the results of the numerical integration of (17) for estimating the ionization rate of atoms in a model dense neutral gas [approximation (18)] and in a nonideal plasma.

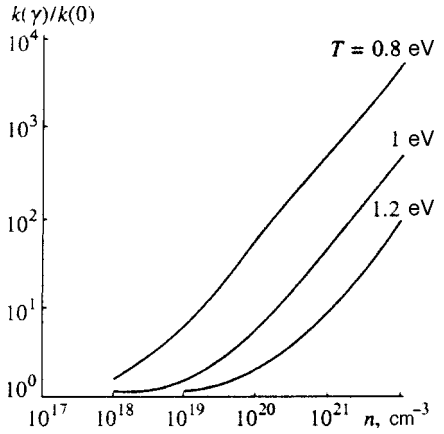

FIG. 3. Rate of the ionization of atoms by electron impact in a model gas with $I = 10$ eV and $\gamma = \text{const}$ as a function of the gas density.

As can be seen from Fig. 3, the ionization rate of a dense gas at low temperatures increases with increasing density according to a power law rather than an exponential law [as is reflected by the second term in (20)]. In Ref. 7 (see also Ref. 8) similar calculations were performed, but with consideration of the ionization potential lowering mechanism, which leads to an exponential dependence of k_{ion} on the gas density, and a comparison with an experiment in liquid Xe was made. Unfortunately, because of the large measurement error at low values of E/N (E is the electric field strength, and N is the atomic density), which specifies the mean electron energy, it is difficult to make an unequivocal choice in favor of a power-law or exponential dependence of k_{ion} on N. The following conclusions can be drawn on the basis of the calculations whose results are presented in Fig. 4: a) the

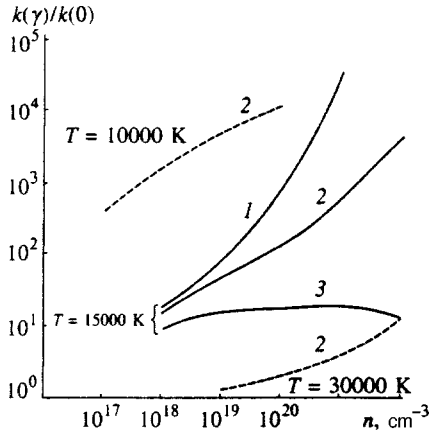

FIG. 4. Ionization rate in a nonideal hydrogen plasma as a function of its density: *1*—ratio of the result of the present work with neglect of the degeneracy of the electrons to the result of the conventional approach; *2*—ratio of the same quantity to the result of the theory in Refs. 1–6; *3*—same as curve *1*, but with consideration of the degeneracy of the electrons.

increase in the ionization rate constant due to consideration of the finite width γ [see (12) and (17)], which introduces power-law tails in the particle momentum distribution, is significantly greater than the increase observed when the ionization potential lowering mechanism is taken into account; b) this acceleration occurs at values of the Coulomb coupling parameter $\Gamma \leqslant 1$, where the theory can be substantiated, while the ionization potential lowering mechanism produces a large effect for $\Gamma > 1$, i.e., outside the region of applicability of the theory; c) consideration of the degeneracy effects is more significant in the theory considered here than in Refs. 1–6. Verification of the theory developed requires special and fairly exact experiments involving, for example, measurements of the excitation or ionization constants in a nonideal plasma or equivalent measurements of the time for establishment of the charge composition.

4. KINETIC EQUATION FOR THE GENERALIZED ELECTRON DISTRIBUTION FUNCTION IN AN ELECTRIC FIELD

The influence of the ionization potential lowering mechanism on the rate of the ionization of atoms by electron impact in a nonequilibrium plasma immersed in a fairly strong electric field was investigated in Refs. 4 and 6. Since, as was shown above, the influence of collisional coupling is not confined to this mechanism alone already under equilibrium conditions, the question of generalizing the theory developed to nonequilibrium conditions arises. Let us consider the model of a Lorentz gas for electrons in an external electric field. This model presumes that the electrons collide only with neutrals, whose distribution function is assumed to be known. The problem thus formulated is linear, and in the case of gases of fairly low density, where the width γ is neglected [i.e., δ_γ in (12) is replaced by an ordinary δ function], it has been thoroughly studied in reference to the description of the kinetics of electrons in weakly ionized gases.[24,25]

The derivation of the generalized kinetic equations is presented after Ref. 18 using the same notation for the electron kinetic Green's functions $G^{\alpha\alpha'}(x_1,x_2)$, where $\alpha,\alpha' = \pm$ correspond to the type of chronological ordering. In the matrix representation with respect to the indices α and α' the electron Green's function satisfies the equations[18]

$$\hat{G}_{01}^{-1}G_{12} = \sigma_z\delta(x_1-x_2) + \int \sigma_z\Sigma_{13}G_{32}d^4x_3, \tag{21}$$

$$\hat{G}_{02}^{-1}G_{12} = \sigma_z\delta(x_1-x_2) + \int G_{13}\Sigma_{32}\sigma_z d^4x_3. \tag{22}$$

Here $\hat{G}_{01}^{-1}$, $\hat{G}_{02}^{-1}$, and σ_z denote the operators

$$\hat{G}_{01}^{-1} = i\frac{\partial}{\partial t_1} + \frac{1}{2m}\Delta_1 - U(x_1) + \mu, \tag{23}$$

$$\hat{G}_{02}^{-1} = -i\frac{\partial}{\partial t_2} + \frac{1}{2m}\Delta_2 - U(x_2) + \mu, \tag{24}$$

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The ordinary kinetic equation is obtained by taking the $(-+)$ components of the matrices G_{12} in Eqs. (21) and (22) and writing down the difference between the terms in these equations:[18]

$$\hat{L}_1 G_{12}^{-+} \equiv (\hat{G}_{02}^{-1} - \hat{G}_{01}^{-1})G_{12}^{-+} = -\int (\Sigma_{13}^{--}G_{32}^{-+} + \Sigma_{13}^{-+}G_{32}^{++} + G_{13}^{-+}\Sigma_{32}^{++} + G_{13}^{--}\Sigma_{32}^{-+})d^4x_3. \tag{25}$$

We introduce new coordinates:

$$t = \frac{t_1+t_2}{2}, \quad \tau = t_1 - t_2, \quad \mathbf{R} = \frac{\mathbf{r}_1+\mathbf{r}_2}{2}, \quad \boldsymbol{\rho} = \mathbf{r}_1 - \mathbf{r}_2.$$

For the operator $\hat{L}_1$ we obtain

$$\hat{L}_1 = -i\left(\frac{\partial}{\partial t} - \frac{i}{m}\nabla_{\mathbf{R}}\nabla_{\boldsymbol{\rho}}\right) + U\left(\mathbf{R}+\frac{\boldsymbol{\rho}}{2}\right) - U\left(\mathbf{R}-\frac{\boldsymbol{\rho}}{2}\right). \tag{26}$$

In the gauge $U(\mathbf{r}_1) = -e\mathbf{F}\cdot\mathbf{r}_1$ the difference between the last terms reduces to

$$U\left(\mathbf{R}+\frac{\boldsymbol{\rho}}{2}\right) - U\left(\mathbf{R}-\frac{\boldsymbol{\rho}}{2}\right) = -e\mathbf{F}\cdot\boldsymbol{\rho}. \tag{27}$$

Going over to Fourier variables and using the definition [compare (1), $f(E,\mathbf{p}) = -(i/2\pi)G^{-+}(E,\mathbf{p},\mathbf{R},t)$]

$$G^{-+}(E,\mathbf{p},\mathbf{R},t) = \int \exp(-i\mathbf{p}\cdot\boldsymbol{\rho} + iE\tau)G^{-+}\left(\mathbf{R}+\frac{\boldsymbol{\rho}}{2},t + \frac{\tau}{2},\mathbf{R}-\frac{\boldsymbol{\rho}}{2},t-\frac{\tau}{2}\right)d^3\rho\, d\tau, \tag{28}$$

after standard transformations[18] we obtain

$$-i\left(\frac{\partial G^{-+}}{\partial t} + \frac{\mathbf{p}}{m}\frac{\partial G^{-+}}{\partial \mathbf{R}}\right) + ie\mathbf{F}\frac{\partial G^{-+}}{\partial \mathbf{p}} = \mathrm{St}_1 G^{-+}. \tag{29}$$

In (29), after Fourier transformation and application of the quasiclassical approximation (compare Ref. 18), the collision integral St_1 reduces to

$$\mathrm{St}_1 G^{-+} = -\Sigma^{-+} G^{+-} + \Sigma^{+-} G^{-+}. \tag{30}$$

Mass operators like Σ^{+-} and Σ^{-+} are depicted in Figs. 1 and 2. In the standard kinetic theory the equation for the particle momentum distribution function is obtained by integrating Eq. (29) over the energy:

$$f(\mathbf{p},\mathbf{R},t) = -\frac{i}{2\pi}\int G^{-+} dE. \tag{31}$$

Here we shall not integrate over the energy, since our problem is to obtain the kinetic equation for a generalized distribution function which depends on the energy and momentum simultaneously.

We note that for $eF\lambda \ll \varepsilon_{\mathbf{p}}$, where λ is the de Broglie wavelength, the retarded (G^R) and advanced (G^A) Green's functions can be assumed to be independent of the electric field F. If the electron density is small and the degeneracy of the electrons can be neglected, the following approximation can be used for the Green's function G^{+-}:

$$G^{+-} = G^R - G^A = -2\pi i \delta_\gamma(E - \varepsilon_{\mathbf{p}}). \tag{32}$$

Using Eqs. (29), (30) and (32) and the explicit forms of the diagrams for Σ^{+-} and Σ^{-+} (of the form in Fig. 1), as well as assuming that the distribution functions for the atoms on which electrons are scattered have the form [compare (7)] $N_a(E,\mathbf{p}) = n(E)\,\delta(E - \varepsilon_{\mathbf{p}})$, we obtain the kinetic equation for the generalized electron energy and momentum distribution function $f(E,\mathbf{p},\mathbf{R},t) = f$:

$$\frac{\partial f}{\partial t} + (\mathbf{v}\cdot\nabla) f - e\mathbf{F}\frac{\partial f}{\partial \mathbf{p}} = \mathrm{St}\{f\}. \tag{33}$$

When elastic and inelastic collisions are taken into account, the linear integral for collisions with atoms whose populations at the energy levels E_i are equal to N_i has the form

$$\mathrm{St}\{f\} = \sum_{i,k} N_i \int d^3q |V_{ik}(\mathbf{q})|^2 \{\delta_\gamma(E - \varepsilon_{\mathbf{p}}) f(E + E_{ki}, \mathbf{p} - \mathbf{q}) - \delta_\gamma(E - E_{ki} - \varepsilon_{\mathbf{p}-\mathbf{q}}) f(E,\mathbf{p})\}. \tag{34}$$

In this model example, for the width γ appearing in the definition (13) of the δ_γ function we can write a nonlinear integral equation (compare Ref. 20) with consideration of only elastic scattering:

$$\gamma = 4\pi^2 \int \frac{d^3p' d^3q}{(2\pi)^6} |V_{\mathbf{q}}|^2 \delta_\gamma(E + \varepsilon_{\mathbf{p}'} - \varepsilon_{\mathbf{p}'+\mathbf{q}}) N_a(\mathbf{p}'). \tag{35}$$

In the simplest approximation the expression for the frequency of transport collisions of electrons with neutrals can be used to estimate γ. It is seen from (33) and (34) that a closed equation for the momentum distribution function can be obtained by integrating these equations over the energy, only if the δ_γ functions are replaced by their limit for $\gamma \to 0$.

Since in the present case consideration of the broadening leads to power-law momentum distributions and to all the effects indicated above of the significant increase in the rates of threshold processes in dense media, when the question of the effects of the deviation from equilibrium is investigated, the kinetic equation (33) presented above should be used under the assumption that E and $\mathbf{p}$ are independent variables. Within the approach in Ref. 6 these effects vanish completely. We note that the generalized distribution function should satisfy the equation obtained by summing Eqs. (21) and (22) to the same accuracy with which the kinetic equation (33) was obtained by finding the difference between Eqs. (21) and (22). Omitting the details of the derivation, we present this equation in the following form:

$$\left\{2(E - \varepsilon_{\mathbf{p}}) - \Sigma^R - \Sigma^A - 2\mu + \frac{1}{4m}\Delta_R - 2e\mathbf{F}\cdot\mathbf{R}\right\} G^{-+} = -\Sigma^{-+}\frac{\{2(E - \varepsilon_{\mathbf{p}}) - \Sigma^R - \Sigma^A - 2\mu\}}{|E - \varepsilon_{\mathbf{p}} - \Sigma^R|^2}. \tag{36}$$

The quantity $\Sigma^R + \Sigma^A = 2\,\mathrm{Re}\,\Sigma^R$ corresponds to twice the density shift of the kinetic energy which leads to the ionization potential lowering mechanism. The quantity

$$\Sigma^{-+} = -\sum_{i,k} N_i \int \frac{d^3q}{(2\pi)^3} |V_{ik}(\mathbf{q})|^2 \delta_\gamma(E - \varepsilon_{\mathbf{p}}) \times f(E + E_{ki}, \mathbf{p} - \mathbf{q})$$

corresponds to the term describing arrival on the right-hand side of Eq. (34). We note that in the homogeneous equilibrium case in the absence of external fields the expression following from (36) for G^{-+} is

$$G^{-+} = -\frac{\Sigma^{-+}}{\Sigma^{-+} - \Sigma^{+-}}(G^R - G^A). \tag{37}$$

It can be shown that this expression is exactly equivalent to Eq. (2). To solve inhomogeneous problems concerning the distribution of electrons with respect to the energy and momentum in an electric field, Eqs. (33) and (36) must be solved together. A similar procedure was used for the generalized radiated intensity $J(\omega,\mathbf{k})$ in Ref. 13, which contains examples of the combined use of kinetic and wave equations that are satisfied by $J(\omega,\mathbf{k})$ [analogs of (33) and (36)] to solve problems concerning the escape of radiation from a finite volume.

In the general case a solution of Eqs. (33), (34), and (36) can be found only by numerical methods, as can the electron energy distribution function in the standard formulation. We shall not dwell here on a more detailed analysis of the kinetic equations obtained. We note only that the numerical solution of the Kadanoff–Baym equations, to which Eqs. (33) are related, has recently been a subject of investigation in describing the kinetics of electrons in semiconductors.[26]

In conclusion, let us formulate the principal results of the present work. It has been shown that in dense media consideration of the frequent collisions which destroy the one-to-one relationship between the energy and the momentum of a particle and lead to the existence of equilibrium power-law tails in the particle momentum distribution function can significantly increase the rate constants of reactions which take place with an appreciable energy threshold. The temperature dependence of such constants can be nonexponential, and they can exhibit a power-law dependence on the density.

Such a non-Arrhenius dependence of the rates of the processes can be manifested in high-pressure chemistry, the breakdown of liquid insulators, the accelerated $V-T$ relaxation in dense molecular gases, and the appearance of hard radiation at relatively low temperatures and high plasma densities.

As a guide, we can point out conditions under which the effects associated with the quantum uncertainty of the energy should occur due to collisions and should be manifested in the processes just cited. The criterion $\hbar\gamma \propto T$ holds at room temperature for gas pressures $\sim 10^3$ atm. As follows from (20), appreciable acceleration of a process in comparison to the classical exponential approximation occurs in the limit $\hbar\gamma/I \gg \exp(-I/T)$. This relation holds for $\hbar\gamma \ll \min(T,I)$, i.e., at far lower pressures. The conditions in a nonideal plasma, under which the effects of the power-law tails strongly influence the rates of the excitation and ionization of atoms and ions, correspond to the criteria $\Gamma \sim 1$ and $\lambda/r_D \sim 1$ and are partially represented in Figs. 3 and 4.

When these effects are verified experimentally, attention should be focused on the fact that the phenomena described here must occur under equilibrium conditions and must not be associated with the presence of a group of fast particles arising from accelerating mechanisms in an unstable plasma.

We sincerely thank V. Yu. Baranov, P. D. Gasparyan, A. M. Dykhne, Yu. M. Kagan, N. G. Koval'skiĭ, A. G. Leonov, A. A. Panteleev, V. D. Pis'mennyi, L. P. Feoktistov, V. E. Fortov, and V. D. Shafranov for their interest in this work and their useful critical comments.

This research was partially supported by the International Scientific-Technical Center (Project No. 076-95).

*)E-mail: award@sunmipt.crec.mipt.ru

[1] M. Schlanges, T. Bornath, and D. Kremp, Phys. Rev. A **38**, 2174 (1988).
[2] T. Bornath, M. Schlanges, and D. Kremp, Contrib. Plasma Phys. **28**, 57 (1988).
[3] Yu. Klimontovich, M. Schlanges, and T. Bornath, Contrib. Plasma Phys. **30**, 349 (1990).
[4] M. Schlanges, T. Bornath, and D. Kremp, in *Physics of Nonideal Plasmas*, W. Ebeling, A. Forster, and R. Radtke (eds.), B. G. Teubner Verlagsgesellschaft, Stuttgart–Leipzig, (1992), p. 154.
[5] U. Leonhardt, in *Physics of Nonideal Plasmas*, W. Ebeling, A. Forster, and R. Radtke (eds.), B. G. Teubner Verlagsgesellschaft, Stuttgart–Leipzig, (1992), p. 162.
[6] D. Kremp, K. Morawetz, M. Schlanges, *et al.*, Phys. Rev. E **47**, 635 (1993).
[7] V. M. Atrazhev, I. T. Iakubov, and V. I. Roldughin, J. Phys. D: Appl. Phys. **9**, 1735 (1976).
[8] A. G. Khrapak and I. T. Yakubov, *Electrons in Dense Gases and Plasma* [in Russian], Nauka, Moscow (1981).
[9] Y. Sakai, in *Linking the Gaseous and Condensed Phases of Matter. The Behavior of Slow Electrons. NATO ASI Series, Ser. B: Physics, Vol. 326*, L. G. Christophorou, E. Illenberger, and W. F. Schmidt (eds.), Plenum, New York (1994), p. 303.
[10] L. G. Christophorou and K. Siomos, in *Electron-Molecule Interactions and Their Applications*, L. G. Christophorou (ed.), Academic Press, New York (1984), Vol. 1, p. 477.
[11] W.-D. Kraeft, D. Kremp, W. Ebeling, and G. Röpke, *Quantum Statistics of Charged Particle Systems*, Plenum Press, New York (1986) [Russ. transl., Mir, Moscow (1988)].
[12] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics: Green's Function Methods in Equilibrium and Nonequilibrium Problems*, Benjamin, New York (1962) [Russ. transl., Mir, Moscow (1964)].
[13] Yu. K. Zemtsov, A. Yu. Sechin, and A. N. Starostin, Zh. Éksp. Teor. Fiz. **110**, 1654 (1996) [JETP **83**, 909 (1996)].
[14] V. M. Galitskiĭ and V. V. Yakimets, Zh. Éksp. Teor. Fiz. **51**, 957 (1966) [Sov. Phys. JETP **24**, 637 (1967)].
[15] Yu. K. Zemtsov, A. Yu. Sechin, A. N. Starostin *et al.*, JETP Lett. **65**, 13 (1997).
[16] Yu. K. Zemtsov, A. Yu. Sechin, A. N. Starostin *et al.*, JETP Lett. **65**, 839 (1997).
[17] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1964)].
[18] E. M. Lifshitz, L. P. Pitaevskii, and L. D. Landau, *Physical Kinetics*, Pergamon Press, Oxford (1981).
[19] I. B. Levinson, Zh. Éksp. Teor. Fiz. **75**, 234 (1978) [Sov. Phys. JETP **48**, 117 (1978)].
[20] Yu. K. Zemtsov and A. N. Starostin, Zh. Éksp. Teor. Fiz. **103**, 345 (1993) [JETP **76**, 186 (1993)].
[21] V. M. Galitskiĭ, *Selected Works. Studies in Theoretical Physics* [in Russian], Nauka, Moscow (1983).
[22] 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).
[23] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[24] N. L. Aleksandrov and É. E. Son, in *Plasma Chemistry* [in Russian], B. M. Smirnov (ed.), Atomizdat, Moscow (1980), No. 7, p. 35.
[25] L. G. H. Huxley and R. W. Crompton, *The Diffusion and Drift of Electrons in Gases*, Wiley, New York (1974) [Russ. transl., Mir, Moscow (1977)].
[26] M. Bonitz and D. Kremp, in *Strongly Coupled Coulomb Systems*, G. J. Kalman, K. B. Blagoev, and J. M. Rommel (eds.), Plenum Press, New York (1998)

Translated by P. Shelnitz

Inhomogeneous two-dimensional structures in liquid crystals

A. N. Bogdanov*) and A. A. Shestakov

Donetsk Physicotechnical Institute, National Academy of Sciences of Ukraine, 340114 Donetsk, Ukraine
(Submitted 15 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 1675–1697 (May 1998)

Localized axisymmetric inhomogeneous states with a continuous distribution of the director field can exist in nematics. Such structures are compressed into dense filaments under the influence of a magnetic or electric field. It is hypothesized that the given states can be achieved in filamentary nematic textures. This model is an alternative to the conventional disclination model. Two types of lattices of axial structures can exist in the entire range of existence of the modulated state. Axial structures with a kernel of finite radius can exist in cylindrical capillaries. The structure and equilibrium dimensions of the axial states are easily altered over a wide range under the influence of an applied field. The feasibility of utilizing isolated axial structures and lattices of such structures in optical data processing and imaging devices is discussed. The most promising outlook in this regard is for modulated states and axial structures in chiral liquid crystals exhibiting spontaneous polarization. © *1998 American Institute of Physics.* [S1063-7761(98)00905-6]

1. INTRODUCTION

Liquid crystals usually form complex textures containing various combinations of point, line, and planar inhomogeneities, which are observed as filaments, loops, inversion walls, nuclei, drops, confocal domains, etc. The appearance of most liquid-crystal textures is determined mainly by the presence of various line defects.[1–3] According to conventional notions, such defects (disclinations) comprise singular lines, in the vicinity of which the field of molecular orientations (i.e., the field of the director $\mathbf{n}(\mathbf{r})$) acquires a discontinuous character. The earliest theory was developed for planar disclinations, i.e., the topic was a structure in which the vector $\mathbf{n}$ rotated in the plane perpendicular to the axis of the disclination. A theory of planar disclinations was formulated by C. W. Oseen and F. C. Frank (see Refs. 1 and 2) for the case of equal elastic constants and was later generalized to the case of arbitrary constants.[4] It is important to emphasize that the results of the theory of planar disclinations have created a model base, which has subsequently been used to describe, classify, and analyze the observed line defects in liquid-crystal textures. Exhaustive empirical material from studies of inhomogeneous states in liquid crystals, spanning a period of several decades, has been organized within the framework of the given model (more information is available in books and survey papers[1–3,5–8]).

By the early seventies, however, it was noticed that the majority of observable line defects in liquid crystals did not conform to the model of planar disclinations. It has been shown theoretically[9,10] that planar disclinations with Frank index $n=2$ in nematics become unstable when the vector $\mathbf{n}$ deviates from the basal plane; the ''digression'' of a disclination into the third dimension forms an axial structure with a continuous distribution $\mathbf{n}(\mathbf{r})$ on the axis. Structures of this kind have been observed in capillaries[10,11] and have also been identified in several common nematic textures.[10] It has subsequently come to light that planar disclinations with an even Frank index are unstable for all realistic parameters in nematics and cholesterics (see Refs. 1 and 12), and only for a definite relation between the elastic constants is it possible for planar disclinations with index $n=\pm 1$ to be stable.[12] In particular, it has been found that only bulk disclinations can exist in such model nematics as *p*-azoxyanisol (PAA) and *n*-methoxybenzylidene *p*–*n*-butylaniline (MBBA).[12] Consequently, the true structure of line defects in liquid crystals differs significantly from the planar disclination model.

In planar disclinations the solutions possess divergent energies both at the center and far from the axis.[1,2] Axial structures with a nonsingular nucleus studied in Refs. 10 and 11 have finite energy on axis, but they are nevertheless delocalized. As we show in the present paper, the equations of the phenomenological theory of liquid crystals also have localized nonsingular solutions with axial symmetry. Similar localized time-independent inhomogeneities in magnetic materials, superconductors, and other branches of nonlinear physics are called *vortices* (see, e.g., Ref. 13). Since the term *vortex* refers to a well-known dynamical construct in liquid crystals, the static two-dimensional axial inhomogeneities in liquid crystals that we study here will be called *axial structures.* To refer to localized line defects with a nonsingular core we can use the term *filament*.

Another type of axial structure is observed in cholesterics with positive anisotropy in electric and magnetic fields parallel to the helicoid axis. Here a square lattice is formed in a certain range of fields.[14–18] Two-dimensional lattices can also be induced in cholesterics by mechanical deformations and temperature gradients.[17]

A bona fide theoretical analysis of axial structures has not been published to date. (Scattered results have been obtained for nematic nonlocalized axial structures in capillaries,[10,11] and the feasibility of localized states has been proved.[19]) In this paper we systematically develop a theory

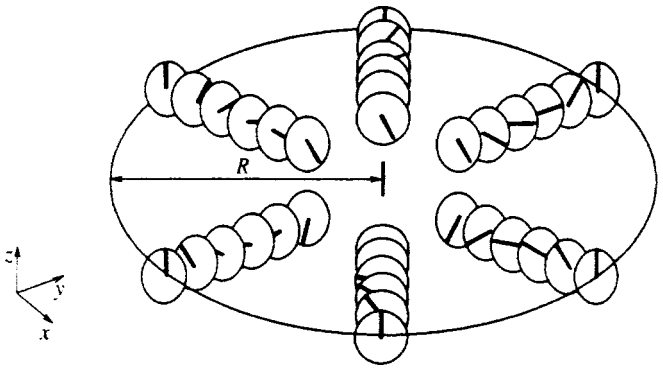

FIG. 1. Distribution of the director in axial structures with $\psi=\varphi+\pi/2$.

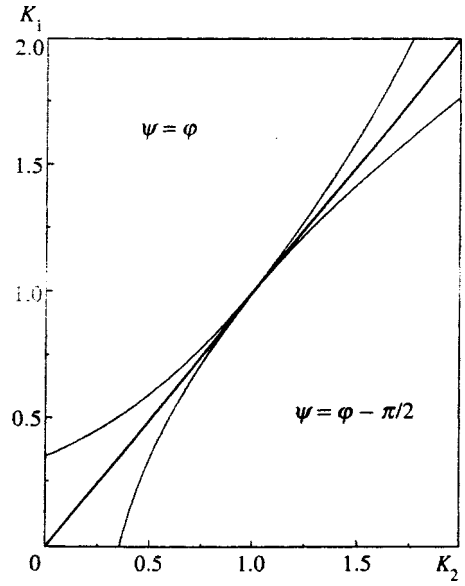

FIG. 2. Phase diagram of axial states in a nematic.

of such states and analyze their properties. We determine the structure and equilibrium parameters of solitary axial formations as well as lattices of such formations in nematic and cholesteric liquid crystals. We investigate the evolution of the structures in external fields, analyze their stability, and determine the boundaries of their existence.

2. NEMATICS

We consider an axisymmetric distribution of the director; on the symmetry axis (Z axis) the vector $\mathbf{n}(\mathbf{r})$ is parallel to this axis and, by a continuous change of orientation of the director, ends up in antiparallel orientation at a large distance from the axis (localized structure, or filament) or attains a certain orientation $\mathbf{n}_0$ at a distance R from the center (nonlocalized structure; see Fig. 1). In contrast to disclinations, these structures are nonsingular inhomogeneities in the bulk of the liquid crystal.

In the continuum theory the possible distributions $\mathbf{n}(\mathbf{r})$ are determined by minimizing the elastic energy

$$W=\int w\,dx=\int \frac{1}{2}[K_1(\mathrm{div}\ \mathbf{n})^2+K_2(\mathbf{n}\cdot\mathrm{curl}\ \mathbf{n}+q_0)^2 + K_3(\mathbf{n}\cdot\mathrm{curl}\ \mathbf{n})^2-\Delta\chi(\mathbf{n}\cdot\mathbf{H})^2]\,d\mathbf{x}, \tag{1}$$

where K_i and q_0 are elastic constants; the last term in Eq. (1) describes the energy density of interaction with an external magnetic field $\mathbf{H}$ (the energy density of interaction with an electric field $\mathbf{E}$ is equal to $\Delta\varepsilon(\mathbf{n}\cdot\mathbf{E})^2$, where $\Delta\chi$ and $\Delta\varepsilon$ are the diamagnetic and dielectric anisotropies, respectively).[1] In the model discussed here (excluding electrical conduction effects) the interaction with a static magnetic field and a static electric field is described in the energy of the system by contributions having the same functional dependence on the material parameters, so that both the magnetic and the electric field lead to analogous orienting effects in the given structures. It is therefore sufficient to limit discussion to the influence of the magnetic field, with the tacit understanding that analogous results hold in an electric field.

The functional (1) describes the energy of a nematic for $q_0=0$ and the energy of a cholesteric for $q_0\neq 0$. The absolute minimum of the system corresponds to solutions that form the potential (1) at zero. In zero field these solutions give a homogeneous state for a nematic and a helicoid structure for a cholesteric.[1] It is clear that all other distributions $\mathbf{n}(\mathbf{r})$ minimizing the energy functional can correspond only to local minima, i.e., they describe metastable states.

We now show that among the metastable states of the system are solutions that describe axial structures. Owing to the axial symmetry of the problem, it is convenient to represent the vector in spherical coordinates, $\mathbf{n}=(\sin\theta\cos\psi,\ \sin\theta\sin\psi,\ \cos\theta)$, and the spatial position variable in cylindrical coordinates, $\mathbf{r}=(\rho,\varphi,z)$. In these new coordinates the energy density assumes the form

$$\begin{aligned} w=&\frac{1}{2}K_1\Big[\cos\ \theta\ \cos(\psi-\varphi)\,\theta_\rho-\sin\ \theta\ \sin(\psi-\varphi)\,\psi_\rho \\ &+\frac{\cos\ \theta\ \sin(\psi-\varphi)}{\rho}\,\theta_\varphi+\frac{\sin\ \theta\ \cos(\psi-\varphi)}{\rho}\,\psi_\varphi\Big]^2 \\ &+\frac{1}{2}K_2\Big[\sin(\psi-\varphi)\,\theta_\rho\ -\sin\ \theta\ \cos\ \theta\ \cos(\psi-\varphi)\,\psi_\rho \\ &-\frac{\cos(\psi-\varphi)}{\rho}\,\theta_\varphi+\frac{\sin\ \theta\ \cos\ \theta\ \sin(\psi-\varphi)}{\rho}\,\psi_\varphi \\ &+q_0\Big]^2+\frac{1}{2}K_3\Big\{\sin^2\theta\Big[\cos(\psi-\varphi)\,\theta_\rho \\ &+\frac{\sin(\psi-\varphi)}{\rho}\,\theta_\varphi\Big]^2+\sin^4\theta\Big[\cos(\psi-\varphi)\,\psi_\rho \\ &+\frac{\sin(\psi-\varphi)}{\rho}\,\psi_\varphi\Big]^2\Big\}-\frac{1}{2}\Delta\chi H^2\ \cos^2\theta. \end{aligned} \tag{2}$$

In Eq. (2) the magnetic field is assumed to be directed along the symmetry (Z) axis of the structure.

We begin the investigation with nematic crystals ($q_0=0$) in zero field. We consider localized solutions. If the vector $\mathbf{n}$ is parallel to Z on the filament axis ($\rho=0$) and becomes antiparallel far from it, the system of equations minimizing the potential (2) has axisymmetric solutions of

the type $\theta=\theta(\rho)$ and $\psi=\psi(\varphi)=\varphi+\gamma$ (where γ is a constant angle). An analysis of the functional (2) for the given class of solutions shows that, depending on the relation between the elastic constants, a stable state of the system corresponds either to the solution $\psi=\varphi+\pi/2$ (I) ($K_1>K_2$) or to the solution $\psi=\varphi$ (II) ($K_1<K_2$). In a narrow neighborhood of the phase equilibrium line $K_1=K_2$ (Fig. 2) the alternative solutions are metastable, i.e., the transition between these structures is of the nature of a first-order phase transition. The standard procedure for calculating the boundaries of stability of extremals yields the following expression for the instability line of the solution (I):

$$\frac{K_2}{K_3}=\begin{cases}\dfrac{-(2t-1)(4t-3)\sqrt{t(1-t)}+[4(2t-1)(1-t)+1]\arcsin\sqrt{1-t}}{-(2t-1)\sqrt{t(1-t)}+[8t(1-t)+1]\arcsin\sqrt{1-t}}, & t=\dfrac{K_1}{K_3}<1\\ 1, & t=1\\ \dfrac{-(2t-1)(4t-3)\sqrt{t(1-t)}+[4(2t-1)(1-t)+1]\ln(\sqrt{t}+\sqrt{1-t})}{-(2t-1)\sqrt{t(1-t)}+[8t(1-t)+1]\ln(\sqrt{t}+\sqrt{1-t})}, & t>1.\end{cases} \tag{3}$$

For the solution (II) the instability line is obtained from Eq. (3) by the substitution $K_1\leftrightarrow K_2$. In Fig. 2 the phase equilibrium line is shown bold, and the boundaries of stability of the individual phases are shown as dotted lines. When $K_1=K_2=K_3$, the solutions become azimuthally degenerate, i.e., the energy of the axial structures does not depend on the angle γ. In this case the functional (1) coincides with the energy functional of an isotropic ferromagnet in zero field.[20]

For all known nematics $K_1>K_2$ (Refs. 1–3), and structures with $\psi=\varphi+\pi/2$, corresponding to zero divergence, are therefore stable in them. We limit the present study to solutions of this kind.

Substituting $\psi=\varphi+\pi/2$ into Eq. (1) and integrating the energy over φ and z, we obtain for the structure (I)

$$W=\pi LK_2\int_0^{\rho_1}\left[\left(\frac{d\theta}{d\rho}+\frac{\sin\theta\cos\theta}{\rho}\right)^2+k\frac{\sin^4\theta}{\rho^2}\right]\rho\,d\rho, \tag{4}$$

where $k=K_3/K_2$ is the ratio of the elastic constants in longitudinal bending (K_3) and torsion (K_2), and L is the length of the filament. In Eq. (4) it is assumed that $\theta(0)=0$ on the filament axis, and the vector **n** becomes antiparallel at a distance ρ_1 from the axis: $\theta(\rho_1)=\pi$. In particular, for $\rho_1=\infty$ Eq. (4) corresponds to the energy of a solitary filament.

The Euler equation for the functional (4))

$$\frac{d^2\theta}{d\rho^2}+\frac{1}{\rho}\frac{d\theta}{d\rho}-\frac{\sin 4\theta}{4\rho^2}-k\frac{\sin^2\theta\sin 2\theta}{\rho^2}=0 \tag{5}$$

in combination with the boundary conditions $\theta(0)=0$ and $\theta(\rho_1)=\pi$ describes the distribution $\mathbf{n}(\mathbf{r})$ in the filament.

In particular, for isolated filaments ($\rho_1=\infty$) in a nematic the multiplication of Eq. (3) by $\rho^2(d\theta/d\rho)$ and integration over ρ yield the first integral

$$\left(\rho\frac{d\theta}{d\rho}\right)^2=\sin^2\theta[1+(k-1)\sin^2\theta]. \tag{6}$$

Separating variables and performing the integration in (6), we obtain the following equation for $\theta(\rho)$:

$$\coth\theta=\frac{1-k(\rho-\rho_0)^2}{2(\rho/\rho_0)}, \tag{7}$$

where ρ_0 is an arbitrary positive number characterizing the dimensions of the core of the filament. At small ρ the angle $\theta\sim\rho$; in the limit $\rho\to\infty$ the angular variable goes as $\theta\sim\pi-1/\rho$. We note that Eq. (7) contains the arbitrary positive number ρ_0. The presence of this free parameter in the solution means that Eq. (7) describes an entire family of solutions of Eq. (5). The significance of this condition will be discussed below. In the one-constant approximation ($k=1$) Eq. (7) coincides with the solution for a vortex in an isotropic ferromagnet[20]:

$$\tan\left(\frac{\theta}{2}\right)=\frac{\rho}{\rho_0}. \tag{8}$$

In the limit $k\to 0$ a singularity is formed in the region $\theta\approx\pi/2$, and for $k=0$ relation (7) goes over to the equation

$$\tan\theta=2\left(\frac{\rho}{\rho_0}\right), \tag{9}$$

which describes the structure with **n** rotated through $\pi/2$.

Figure 3 shows the profile of $\theta(\rho/\rho_0)$ (7) for various

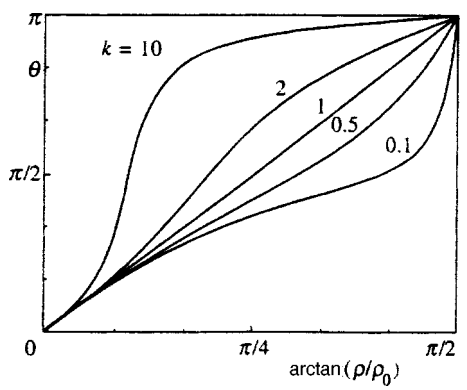

FIG. 3. Axial structures in a nematic for various values of k.

values of k. For $k<1$ the $\theta(\rho/\rho_0)$ curves have a characteristic hump shape; for $k>1$ the profile (7) contains an inflection point.

Substituting the solutions (7) into (4) and integrating, we obtain expressions for the energy of isolated filaments:

$$\frac{E_V}{2\pi K_2 L}=\begin{cases}1+\dfrac{k}{2\sqrt{1-k}}\ \ln\dfrac{1+\sqrt{1-k}}{1-\sqrt{1-k}}, & 0<k<1\\ 1+\dfrac{k}{\sqrt{1-k}}\ \arcsin\sqrt{\dfrac{k-1}{k}}, & k>1\end{cases}. \tag{10}$$

The energy (10) has the minimum value $E_V=2\pi K_2 L$ for $k=0$ and increases monotonically as k increases. For $k=1$ the filament energy is $4\pi K_2 L$ and coincides with the energy of a vortex in an isotropic ferromagnet.[20] At large k the filament energy is proportional to $\sqrt{K_2 K_3}$.

For fixed k the solutions (7) represent a family of functions of the type $\theta(\rho/\rho_0)$. The existence of solutions of the type (7) is a consequence of the invariance of the nematic energy (2) under radial extension ($\rho_0>1$) or compression ($0<\rho_0<1$) of the functions $\theta(\rho)$. Consequently, the resulting vortex states in a nematic have an ''amorphous'' character: For a fixed value of k the energy of the filaments (10) remains constant under arbitrary radial compression or extension of the profile $\theta(\rho)$.

Equations (6) and (7) can be used to determine the functional relation between $d\theta/d\rho$ and θ:

$$\frac{d\theta}{d\rho}=f(\theta)=\frac{1}{\rho_0}\sqrt{1+(k-1)\sin^2\theta}\,(\cos\theta+\sqrt{1+(k-1)\sin^2\theta}). \tag{11}$$

The phase portraits (11) of the solutions (7) begin at the points $(0, 2/\rho_0)$ and end at the point $(\pi,0)$. The trajectories (11) fill the entire phase plane $(\theta, d\theta/d\rho)$ as ρ_0 varies from zero to infinity.

Besides the localized solutions (7) for isolated filaments, Eq. (5) also has axisymmetric solutions with the boundary conditions

$$\theta(0)=0, \quad \theta(\rho_1)=\theta_0 (0<\theta_0<\pi). \tag{12}$$

From the form of the solutions (7) it follows that a function passing through the point (ρ_1,θ_0) can be found in the set of solutions $\theta(\rho/\rho_0)$ (7) by an appropriate choice of ρ_0. It is clear that the resulting profile will describe the solutions of Eq. (5), subject to the boundary conditions (12).

Axial structures of finite radius can be induced in capillaries. The boundary conditions (12) hold in capillaries of radius ρ_1 when suitable orienting films are sprayed onto the walls.[1–3] Axial states of this kind with $\theta_0=90°$ have been observed previously.[10,11]

The indicated properties of structures with $\psi=\varphi+\pi/2$ remain in effect for the solutions $\psi=\varphi$. Because so far there are no known substances in which defects of the given type can be stable, we omit any discussion of the properties of these solutions. The influence of magnetic and electric fields on the investigated nematic structures will be discussed in Sec. 3.3.

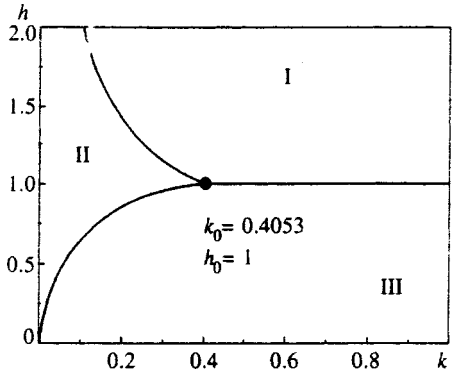

FIG. 4. Phase diagram of helicoid states in a cholesteric. I) Homogeneous state; II) conical phase; III) cycloid.

3. CHOLESTERICS

The ground state in cholesterics corresponds to a one-dimensional modulated structure — a helicoid. In cholesterics with positive anisotropy of the susceptibility ($\Delta\chi>0$) in a magnetic field perpendicular to the helicoid axis and in cholesterics with $\Delta\chi<0$ in a field parallel to the helicoid axis the helical structure is distorted by the field (the field-distorted helicoid is called a cycloid), and the period of the structure increases as the field is increased; finally, in the critical field

$$H_c^2=\pi^2 q_0^2 K_2/4\Delta\chi, \tag{13}$$

unbounded growth of the period takes the system into a homogeneous state.[21,22] Experimental studies of the evolution of a cycloid in magnetic and electric fields have been reported for a great many cholesterics.[23–26]

In this section we need to know the boundaries of existence and energies of various helical structures in a helicoid. We therefore summarize the results of calculations of the equilibrium parameters of helicoid states in Refs. 1–3, along with those obtained in the present study (Fig. 4). In analyzing the equations for various helical structures in cholesterics, we have found that in the range of small k [$k<k_c=(2/\pi)^2=0.4053$, where the parameter k is defined in Eq. (4)] in a certain critical field H_1 (Fig. 4) the axis of the helix rotates parallel to the magnetic field, transforming the helicoid into a conical phase. Subsequent evolution of the system takes place by virtue of the decreasing angle of the cone (the vector **n** bends toward the direction of the field); finally, in the upper critical field $H_2=(2\pi\sqrt{k})H_c$ the cone collapses, and we have a transition to a homogeneous state. Regions of thermodynamic stability of the individual phases are mapped in the $H-k$ phase diagram (Fig. 4).

We now show that together with one-dimensional modulated structure (various helicoids), two-dimensional modulated structures and two-dimensional localized states also enter into the set of metastable solutions of the equations of state of a cholesteric.

We now consider axisymmetric structures in cholesterics in a field directed along the symmetry axis. The minimum of the functional (2) in this case corresponds to structures with $\psi=\varphi+\pi/2$. Integrating the functional (2) over φ and z and introducing the new length and field units

$$r=\frac{\rho}{\rho_0}, \quad h=\frac{H}{H_c}, \quad \rho_0=\frac{\pi}{q_0} \tag{14}$$

(H_c is given by Eq. (13)), we can write the energy of the cholesteric in the form

$$w=\frac{W-W_0}{\pi L K_2}=\int_0^R \tilde{w} r\, dr$$
$$=\int_0^R\left[\left(\frac{d\theta}{dr}+\frac{\sin\theta\cos\theta}{r}-\pi\right)^2 + k\frac{\sin^4\theta}{r^2}+\frac{\pi^4}{4}h^2\sin^2\theta-\pi^2\right] r\, dr, \tag{15}$$

where $W_0=\pi^3 L K q_0^2(1-h^2/4)$ is the energy of the homogeneous state with $\mathbf{n}\|Z$, and L and k are defined in Eq. (4).

The Euler equation for the functional (15)

$$\frac{d^2\theta}{dr^2}+\frac{1}{r}\frac{d\theta}{dr}+\frac{\sin 4\theta}{4r^2}-k\frac{\sin^2\theta\sin 2\theta}{r^2}+2\pi\frac{\sin^2\theta}{r} - \frac{\pi^4 h^2}{8}\sin 2\theta=0 \tag{16}$$

with the boundary conditions

$$\theta(0)=0, \quad \theta(R)=\pi \tag{17}$$

describes an axial structure of radius R in a cholesteric. As in nematics, states with a finite radius can occur in cylindrical capillaries, and isolated states in the bulk of a cholesteric correspond to the boundary condition $R=\infty$.

For $k=1$ Eq. (16) coincides functionally with the equation for two-dimensional axisymmetric distributions of the magnetization (''vortices'') in uniaxial ferromagnets without an inversion center.[27–29]

Unlike Eq. (5), this equation cannot be solved analytically. A standard difference method is suitable for the numerical integration of Eq. (16) with the boundary conditions (17). After linearization in the vicinity of the trial function the differential equation reduces to a system of linear algebraic equations for the values of the function $\theta(r)$ at discrete sites. The interaction matrix for the given system has a band structure: only elements on the principal diagonal and a few ''parallels'' are nonvanishing. Special computational methods and procedures described in Ref. 30 have been used to solve such systems of equations.

However, the convergence of the differencing method is largely determined by an auspicious choice of trial function. To look for possible solutions of the boundary-value problem (16) and (17) and to determine their nature, it is necessary to switch to an auxiliary Cauchy problem, specifically to begin by solving Eq. (15), subject to the initial condition

$$\theta(0)=0, \quad \left(\frac{d\theta}{dr}\right)_{r=0}=a\,(0\leqslant a\leqslant\infty). \tag{18}$$

The numerical calculation of the functions $\theta(r)$ (in this paper we use the Runge–Kutta method) for various values of a provides a means of analyzing the behavior of the integral curves of the Cauchy problem and choosing those which satisfy the given boundary-value problem (or proving that none exists). The possible integral curves of Eqs. (16) and (18) have been analyzed on the phase planes (r,θ) and $(\theta, d\theta/dr)$. Whenever solutions corresponding to the boundary-value problem could be found, the profiles obtained by the Runge–Kutta method served as trial functions for the differencing method.

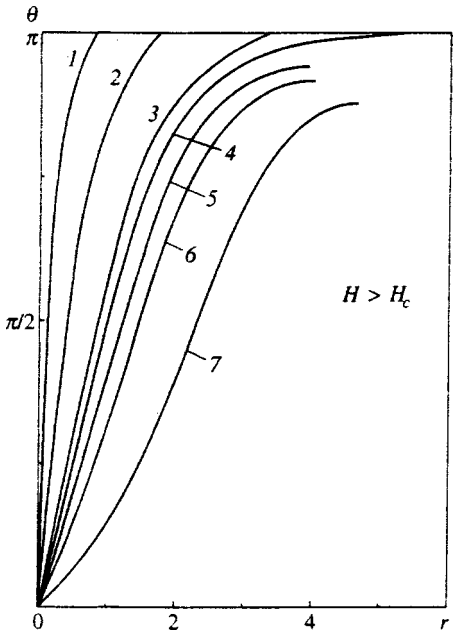

FIG. 5. Behavior of the solutions of the boundary-value problem for axial structures in a cholesteric for $H>H_c$.

The numerical calculations and analytic investigation have led to the following results.

A. Localized Structures and Solutions with Finite Radius

The nature of the solutions of Eq. (16) varies significantly according to whether the applied magnetic field is above or below the critical level H_c (13).

In supercritical fields, $H>H_c$, the energy of inhomogeneous structures in a cholesteric is higher than the energy of the homogeneous state. The inhomogeneities therefore tend to localize here. The $\theta(r)$ profiles for the solutions of the boundary-value problem (16) and (17) are shown in Fig. 5, and the corresponding phase portraits are shown in Fig. 6. The behavior of $\theta(r)$ depends on the value of the parameter a (18). If a is lower than a critical value $a_0(h,k)$, the solutions $\theta(r)$ describe structures with finite radii $R(a)$ and a nonzero derivative at boundary points $((d\theta/dr)_R>0)$ (curves *1*, *2*, and *3* in Fig. 5). For $a<a_0$ the trajectories of the phase portraits for the integral curves of the Cauchy problem (I) cross the axis $\theta=\pi$ for $d\theta/dr>0$ and terminate at one of the poles $(\pi(2n+1)/2, 0)$, $n=1, 2, 3, \ldots$ (the phase trajectory with a pole at the point $(3\pi/2, 0)$ is shown in Fig. 6). Along these curves the solutions of the boundary-value problem (17) and (18) correspond to the range $0\leqslant\theta\leqslant\pi$ (Fig. 6). The boundary conditions for such solutions can be established in cylindrical capillaries by appropriate surface treatment.

For $a>a_0$ the functions $\theta(r)$ describe solutions of finite radius R_0 with the vector $\mathbf{n}$ rotated through an angle smaller

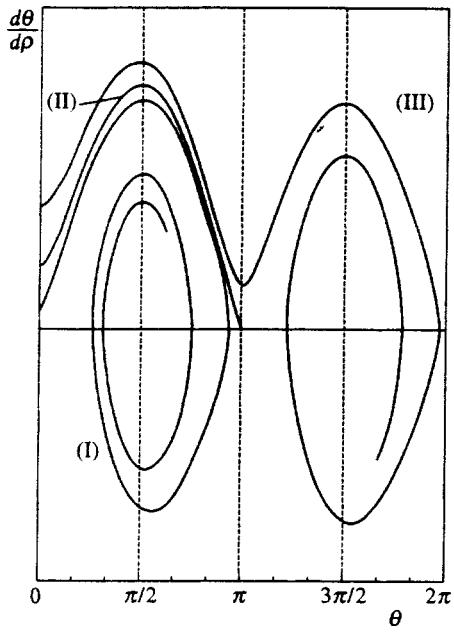

FIG. 6. Phase portraits of the solutions shown in Fig. 5.

than π, where $(d\theta/dr)=0$ at $r=R_0$ (curves *5*, *6*, and *7* in Fig. 5). In Fig. 6 these solutions correspond to segments of trajectories terminating at the pole $(\pi/2, 0)$ (III).

Finally, for $a=a_0$ we encounter localized solutions describing isolated filaments in the bulk of the cholesteric (curve *4* in Fig. 5). These solutions correspond to separatrix trajectories of the type (II) beginning at the vertical axis and terminating at the point $(\pi,0)$.

Figure 7 shows the profiles of localized solutions for $k=4$ and several supercritical values of the field. Localization of the filaments intensifies drastically as the field is increased. The effective dimensions of the filament core decrease monotonically with increasing field (the solutions do not become unstable in this case). The process naturally culminates with the filament cores being compressed to such a degree that the distances between molecules become commensurate with their diameters. The methods of the continuum theory are no longer valid in this region.

As the field decreases in the range $H>H_c$ and approach the critical value, the filaments expand, and their energy approaches the energy of the homogeneous state. A narrow transition region begins to form between the core with $\theta\approx 0$ and the outer region, or ''domain boundary.'' As $H\to H_c$, unbounded expansion of the core takes the filament into a homogeneous state. Localized axial structures do not exist in cholesterics for $H\leqslant H_c$.

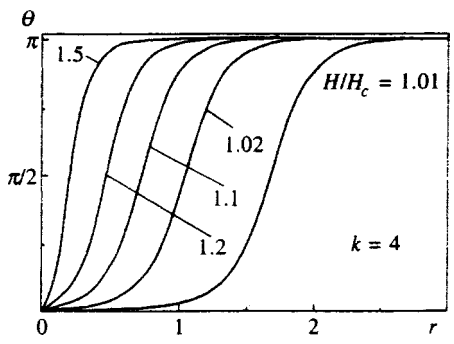

FIG. 7. Localized axial structures of a cholesteric for various values of the magnetic field, $k=4$.

In the range $0<H<H_c$ the energy of the modulated states is lower than the energy of the homogeneous phase, so that the solutions for axial structures have an altogether different nature here. For arbitrary values of the parameter $a>0$ the solutions of the Cauchy problem (16) and (18) are functions $\tilde{\theta}(r)$ that increase monotonically to infinity. These solutions correspond to a modulated structure consisting of concentric circles (''water ripples''). The solutions of the investigated boundary-value problem on the $\tilde{\theta}(r)$ trajectories correspond to segments $0\leqslant\theta\leqslant\pi$ describing axial structures of finite radius. Small values of the parameter a correspond to axial solutions of large radius. Such $\theta(r)$ profiles are cylindrical domains, whose ample inner region existing in the homogeneous state with $\theta\approx 0$ is separated from the outer region with $\theta=\pi$ by a thin 180° wall. In the limit $R\to\infty$ the diameter of this domain grows without bound, and localized structures are therefore unstable in cholesterics in fields below the critical value H_c (13).

This result has a simple physical interpretation. For $H<H_c$ the energy of the inhomogeneous state has a lower energy than the homogeneous state. Consequently, as the profile expands indefinitely, the volume occupied by the energetically favorable inhomogeneous part (the ''domain boundary'') grows without bound. Thus, upon transition into the region where modulated structures exist, unbounded expansion of the core of the structure must take place in the localized filaments present in the bulk of the cholesteric (in other words, they must ''swell and burst'').

B. Lattices of Axial Structures

Inasmuch as the energy of the inhomogeneous state of a cholesteric is lower than the energy of the homogeneous state in the entire region where $H<H_c$, structures with the maximum possible extent of inhomogeneity should be encountered here. One example of such a state is the helicoid. In this case the vector **n** varies periodically in a certain direction (axis of the helicoid). The helicoid can be regarded as a one-dimensional lattice or a one-dimensional modulated structure. In effect, the inhomogeneity of the structure, and hence the energy stability, increase in the presence of modulation in other directions as well, i.e., in two-dimensional and three-dimensional lattices. We know that the thermodynamic stability of two-dimensional lattices of inhomogeneous states has been proved for superconductors in the mixed state (lattices of Abrikosov vortices) (Refs. 31 and 32) and in a definite range of magnetic fields for magnets without an inversion center (lattices of magnetic vortices) (Ref. 28). Related structures have been observed in cholesterics under certain conditions: square lattices[15–18] and three-dimensional modulated structures (so-called blue phases) (Ref. 33). In this section we discuss two-dimensional lattices of the axial structures obtained above.

The stabilization of modulated structures in cholesterics and in other chiral systems is associated with the energy of Lifshitz invariants in the system, which ensure that the energy of the system will decrease only when the order parameters vary in a definite direction (chirality!).[1] In particular, in cholesterics this term has the form $q_0\mathbf{n}$ curl $\mathbf{n}$ (see (I)), and the sign of the parameter q_0 in a given substance determines the energetically favorable direction of rotation of the vector $\mathbf{n}$ in it. Consequently, of all the various lattices of axial structures in a cholesteric, the only ones that ensure a decrease in the system energy (relative to the homogeneous state) are those in which a specified direction of rotation of the director $\mathbf{n}$ is preserved. This condition is obviously satisfied by lattices in whose unit cell the vector $\mathbf{n}$ is parallel to the symmetry axis (Z axis) and, by 180° rotation, becomes antiparallel to the cell boundaries. We refer to such structures as π lattices. Also possible are lattices with films in which $\mathbf{n}\|Z$ and, by 90° rotation, become perpendicular to the axis of the axial structure at the cell boundaries ($\pi/2$ lattices).

The physical equivalence of states with $\mathbf{n}$ and $-\mathbf{n}$ can be exploited in this case as well to maintain rotational invariance of the director in the lattice. It is obvious that the second type of lattice cannot exist in systems whose states are described by a vector order parameter (e.g., in ferromagnets). It is important to note that in $\pi/2$ lattices consisting of triangular, square, or hexagonal cells the $\mathbf{n}(\mathbf{r})$ field has discontinuities at the vertices. These singularities are radial line disclinations. They can also be removed by the ''digression'' of a disclination into the third dimension. Consequently, in $\pi/2$ lattices with a continuous structure the main lattice is accompanied by the formation of a system of cells with centers at the nodes of the lattice.

The equilibrium states of two-dimensional lattices in a cholesteric are determined by solving the system of differential equations for $\theta(x,y)$ and $\psi(x,y)$, which minimize the energy functional (1). This problem is difficult to solve, even numerically. The problem can be simplified considerably by invoking the approximation of circular cells.[30,28] In this approximation the unit cell of a lattice with a hexagonal (or square) cross section is replaced by a circular cylinder of equal volume. Accordingly, the boundary conditions are replaced by circular counterparts: by the rotation (17) for π cells, and for $\pi/2$ cells by the relation

$$\theta(0)=0, \quad \theta(R)=\pi/2 \tag{19}$$

(it is assumed here that in $\pi/2$ lattices both systems of cells have identical dimensions).

In the given approximation the problem of calculating the equilibrium structure of the lattices restores axial symmetry and reduces to the integration of Eq. (16), subject to the boundary conditions (17) or (19), with subsequent minimization of the lattice energy density

$$F=w/\pi R^2 \tag{20}$$

with respect to R (the energy w is given by Eq. (15)). Consequently, the solutions of Eq. (15) with finite radius correspond to a vortex lattice. As mentioned above (see Figs. 5 and 6), for $H<H_c$ Eq. (16) with the boundary conditions (17) has solutions for arbitrary radii. The corresponding profiles for structures with the boundary conditions (19) can be obtained by drawing the straight line $\theta=\pi/2$ in Fig. 5.

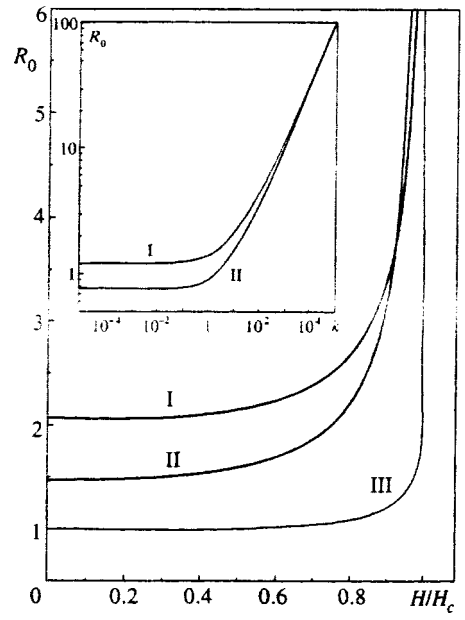

FIG. 8. Equilibrium radii of unit cells. I) π lattice; II) $\pi/2$ lattice; III) helicoid. Inset: R_0 versus k in zero magnetic field. I) π lattice; II) $\pi/2$ lattice.

The procedure for numerically calculating the optimum lattice radii entails solving Eq. (16) for various values of R and comparing their energies (20). The calculations show that the energy (20) has a minimum for finite cell radii R in the entire region $0\leqslant H<H_c$ for both types of lattices. In particular, for π lattices the minimum energy is attained at $R_0=1.37$ for $k=1$ and at $R_0=1.68$ for $k=4$; in the case of $\pi/2$ lattices the minimum occurs at $R_0=1.08$ for $k=1$ and at $R_0=1.31$ for $k=4$. (By way of comparison, the half-period of a helicoid structure in a cholesteric is equal to unity.[3])

Usually in cholesterics the elastic constant in longitudinal bending K_3 is several times the torsion constant K_2, so that the parameter k (4) assumes values of the order of unity. In a mixture of cholesteryl nonanoate with MBBA, for example, Hervet et al.[15] give $k=3.886$ and $T=22°$C. However, in the vicinity of spontaneous phase transitions the parameter $k(T)$ can vary over a wide range. At small k the distribution $\theta(r)$ in equilibrium structures of π lattices is generally linear, and the core increases in size as k increases. A calculation shows that the functions $\theta(\mathbf{r})$ describe the equilibrium profiles of $\pi/2$ lattices to within one percent for $\theta\leqslant\pi/2$.

Graphs of the equilibrium cell radius as a function of the parameter k are shown in Fig. 8 for both types of lattice in zero field. The period of the lattices increases monotonically with k, and the cells of π lattices have larger radii for all values of k.

Over a wide range of fields ($H<H_c$) the equilibrium lattice periods exhibit weak field dependence, but begin to grow without bound as the critical field H_c is approached (Fig. 8). Figure 9 shows the evolution of the equilibrium structure of the unit cells in π lattices as the magnetic field is increased. In the vicinity of the critical field the inhomogeneity in the distribution of $\mathbf{n}$ is localized in a narrow transi-

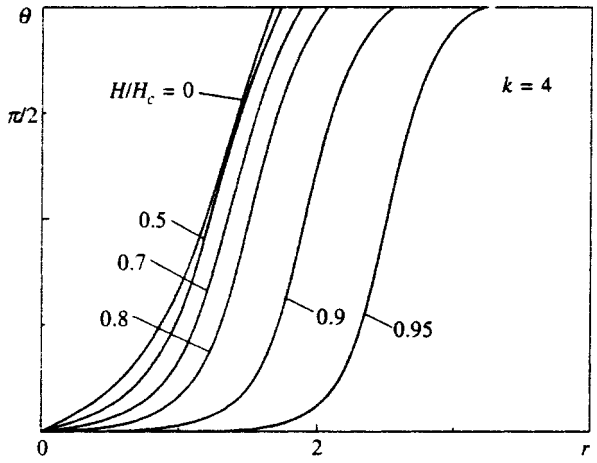

FIG. 9. Equilibrium structures of the unit cells of a π lattice for various values of the field, $k=4$.

tion region. In the cells of $\pi/2$ lattices, conversely, the inhomogeneities are concentrated at the cell boundaries. This difference is readily understood if we recall that states with $\theta=\pi/2$ have the maximum energy in a magnetic field. These states occur at the cells boundaries in $\pi/2$ lattices, whereas in the alternative lattices they occur in the interior of the cells. The inhomogeneities are localized specifically in the region of states close to $\pi/2$ in both cases. As $H\rightarrow H_c$, the cell radii increase sharply in both of the investigated lattice types, but the $\pi/2$ lattices are characterized by a steeper increase in the equilibrium dimensions of the structure (Fig. 8).

Throughout the entire region of existence of modulated structures the energy density of $\pi/2$ lattices is greater than the energy density of π lattices (Fig. 10). Like the helicoid,[21,22] both of the investigated lattice types go over to a homogeneous state by virtue of the unbounded growth of the system period in the field H_c. Clearly, the energies of all modulated structures become equal to the energy of the homogeneous state in this critical field.

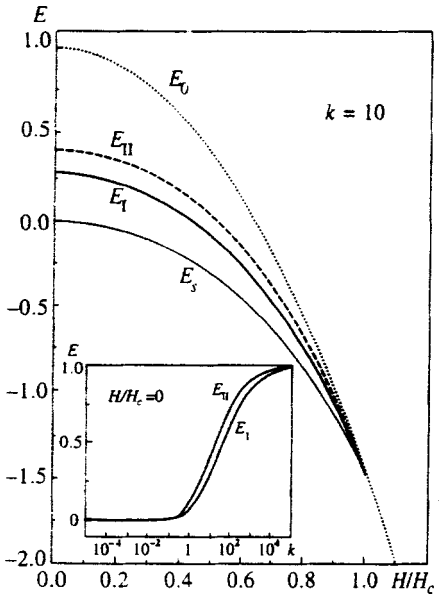

FIG. 10. Equilibrium energies of axial π lattices $E_{\rm II}$ and $\pi/2$ lattices $E_{\rm I}$ versus magnetic field, $k=10$. For comparison the thin curve represents the helicoid energy E_s, and the dotted curve represents the energy of the homogeneous state E_0. Inset: equilibrium energies of axial lattices versus k in zero magnetic field ($E_{\rm II}$ and $E_{\rm I}$ correspond to π and $\pi/2$ lattices, respectively).

The difference between the distributions of the inhomogeneities in the helicoid and in both lattice types disappears as $H\rightarrow H_c$; the modulated states discussed here form a system of growing regions occupied by the homogeneous state with $\theta=0$ (domains), and the inhomogeneities are localized in narrow layers (180° domain boundaries). In this region, therefore, the inhomogeneous states can effectively be described by the model of a solitary thin, plane domain boundary.[30] In the nonequilibrium energy of the system the energy contribution of these inhomogeneous regions can be represented by a global parameter: the energy density of the domain boundaries σ. The energy density σ is negative in the entire region where $H\leqslant H_c$ and is equal to zero in the critical field. The transition to the homogeneous state takes place as a result of the unbounded expansion of the domains and the displacement of the domain boundaries to infinity. Consequently, the nature of the transition of modulated structures into the homogeneous state enables us to determine the boundaries of their existence from an analysis of the plane domain boundary of a cholesteric.

We recall that an analogous situation occurs in superconductors[31] and in magnets without an inversion center.[27–29] For cholesterics the role of the Ginzburg–Landau parameter is played by the quantity h (14), which characterizes the relative contribution of interactions that form equilibrium modulated structures.

We close this section with a discussion of the limits of validity of the results obtained in the circular-cell approximation. It is clear that the structure of the cells of two-dimensional lattices, calculated by exact solution of the appropriate differential equations, preserves axial symmetry only at the center; near the cell boundaries the symmetry of the lattice is manifested, and the distributions $\theta(x,y)$ and $\psi(x,y)$ deviate substantially from axial. An analysis shows that taking the lattice symmetry into account cannot significantly influence the results obtained above. An important consideration here is that the main contribution to the energy of the system is from zones located at the centers of the cells. Consequently, the variation of the structure in the peripheral regions of the cells only slightly affects the overall energy balance of the system. For example, a variation of the radial components in an axial cell of a π lattice does not alter the results by more than two percent.[28]

All the same, it is important to calculate the equilibrium states of the unit cells with allowance for their real symmetry. In zero field the absolute minimum of the functional (1) corresponds to a helicoid. However, the difference between the energies of helicoid structures and two-dimensional modulated structures does not exceed ten percent and diminishes as the field is increased. Consequently, this consideration leaves open the question of which of the modulated structures is thermodynamically stable in a magnetic field.

Upon transition into the homogeneous state, all the modulated phases of a cholesteric are transformed into struc-

tures with highly localized inhomogeneities. In this region the limitations imposed by using the circular-cell approximation are lifted, and the results, including the critical field values, can be regarded as exact.

C. Radial Stability of the Solutions

In the overall energy balance of an axial structure (15) we can distinguish components with different types of dependence on the spatial variable: the regular part of the elastic energy (w_r), terms contained in the invariant **n** curl **n** (w_c), and the energy of interaction with the external field (w_h):

$$w_r=\int_0^R\left[\left(\frac{d\theta}{dr}+\frac{\sin\theta\cos\theta}{r}\right)^2+k\frac{\sin^4\theta}{r^2}\right]r\,dr,$$

$$w_c=-2\pi\int_0^R\left(\frac{d\theta}{dr}+\frac{\sin\theta\cos\theta}{r}\right)r\,dr,$$

$$w_h=\frac{\pi^4}{4}h^2\int_0^R\sin^2 r\theta\,dr. \tag{21}$$

The energy w_r is proportional to $[(d\theta/dr)^2+1/r]r^2\sim r^0$, i.e., it is zero-dimensional in r; $w_c\sim r^1$ is linear, and w_h is proportional to r^2. These differences in the number of dimensions of the energies of different interactions have a significant influence on the character of the solutions for axial states. To corroborate this assertion, we consider the variation of the energy (15) in the scale transformation

$$r=(1+a)r' \tag{22}$$

of a $\theta(r)$ profile of radius R. Performing the integration in Eq. (15), we can write an exact expression for the energy of the deformed $\theta(r')-w'$ profile:

$$w'=w+(w_c+2w_h)a+w_h a^2. \tag{23}$$

Equation (23), first derived for localized two-dimensional structures (vortices) in magnets without an inversion center,[29] is valid for fairly general models[34] and leads to important conclusions about the stability of axial states. It follows from Eq. (23) that a $\theta(r)$ profile with the boundary conditions (16) is stable under the transformation (22) if

$$w_c+2w_h=0,\quad w_h>0. \tag{24}$$

To verify whether the solutions of Eq. (16) satisfy these conditions, we multiply it by $r^2(d\theta/dr)^2$ and integrate from 0 to R. Substituting the boundary conditions at zero, we obtain

$$w_c+2w_h=-[r^2(d\theta/dr)^2]_{r=R}+\left[\cos^2\theta+(k-1)\sin^4\theta+\pi r\sin 2\theta+\frac{\pi^4h^2}{8}\sin^2\theta\right]_{r=R}. \tag{25}$$

For localized solutions of Eq. (16) the right-hand side of Eq. (25) vanishes, and conditions (24) are therefore satisfied.

For solutions of finite radius, according to (17), $(d\theta/dr)_{r=R}>0$. Consequently, here we have $w_c+2w_h<0$ and, according to (23), the given solutions are unstable under radial expansion. For axial solutions, in which $(d\theta/dr)_{r=R}=0$ and $\theta(R)<\pi$ at the boundary, condition (25) has the form $w_c+2w_h>0$. Such structures are therefore unstable under compression. Consequently, only localized states preserve stability under radial perturbations of the type (22) in the bulk of a cholesteric. Other solutions are radially unstable, and expansion or compression transforms their profiles into types corresponding to localized solutions. However, axial structures of finite radius can be induced in cylindrical capillaries.

The energy w_c, which lowers the energy of localized states in cholesterics, plays a decisive role in the stabilization of such states. In centrosymmetric materials we have $w_c=0$, and the condition (24) for the radial stability of axial structures is not satisfied (see Ref. 29 for details).

In this regard we consider the axial solutions (7) for nematics. It follows from Eq. (21) that such solutions are unstable in a magnetic field. Here the energy attains a minimum at $R=0$ in a magnetic field. In all known nematics, $\Delta\chi>0$ and thus $w_h>0$. This means that the investigated axial lines in nematics become compressed under the influence of a magnetic field. The process does not exhibit threshold behavior in the given ideal model: the haphazard influence of even weak magnetic fields (e.g., the earth's magnetic field) can cause the filamentary structures to collapse. This effect results in the formation of dense localized structures, in which the intermolecular distances are comparable to the molecular diameters. The ''dark filaments'' in filamentary nematic textures can be assumed to represent just such formations.

For many years filamentary textures now known as nematics, like other inhomogeneous states in liquid crystals, have been explained on the basis of disclination theory.[1] According to the descriptions, the ''filaments'' attached to support surfaces in filamentary textures are accompanied by a fair number of similar formations with one or two free ends. These line defects are highly mobile: they bend easily and are carried off by a flow. This behavior of the free filaments is consistent with their interpretations as localized formations having a continuous structure. The free filaments can comprise localized structures of finite length terminating in point defects. Disclinations, in turn, are nonlocalized states with a discontinuity of the director field. The motion of such defects requires considerable energy expenditures, because it involves major reorientation of large volumes of the nematic. Careful observations of various defects in filamentary textures should reveal the actual structure of the ''dark filaments.'' We recall that the theoretical prediction of the existence of axial structure with a nonsingular kernel[9,10] was followed by focused investigations, from which it was possible to identify these objects in a number of the most common nematic textures.[10]

Nematic axial structures have also been observed in cylindrical capillaries,[10,11] but the influence of fields on these structure has not been investigated. Since the axial structures did not collapse in the earth's magnetic field in Refs. 10 and 11, we can assume that the compression of axial lines in real materials is masked by effects of interaction with the surface and defects or is possibly a threshold process.

D. Linear Ansatz for Axial Solutions

To analyze the basic laws governing the formation of axial structures in liquid crystals, we consider the energy (15) for the elementary model distribution

$$\theta(r)=\pi\frac{r}{R}(r<R),\quad \theta(r)=0(r>R)\quad (\pi\text{-lattice}), \tag{26}$$

$$\theta(r)=\frac{\pi}{2}\frac{r}{R}(r<R),\quad \theta(r)=0(r>R)\quad (\pi/2\text{-lattice}), \tag{27}$$

The substitution of relations (26) and (27) into the energy (15) and subsequent integration yield the following result for the energy densities of the lattice w (20) with the model distribution (26) and (27):

$$E=\frac{\tilde{w}}{q_0^2}=\frac{A}{R^2}-\frac{B}{R}+C. \tag{28}$$

Here R serves as the parameter to be varied. For π lattices the coefficients A, B, and C have the values

$$A_1=\int_0^1\left[\left(1+\frac{\sin 2\pi x}{2\pi x}\right)^2+k\frac{\sin^4\pi x}{\pi^2x^2}\right]x\,dx=(0.540+0.084k),$$

$$B_1=\int_0^1\left(2x+\frac{\sin 2\pi x}{\pi}\right)dx=1,$$

$$C_1=\frac{\pi^2h^2}{4}\int_0^1(\sin^2\pi x)x\,dx=0.617h^2, \tag{29}$$

and the integration for $\pi/2$ lattices leads to the results

$$A_2=(0.257+0.053k),\quad B_2=0.5(1+4/\pi^2),$$

$$C_2=0.867h^2.$$

The integration of Eq. (15) with the model solution (26) yields the potential

$$F=A-BR+CR^2, \tag{30}$$

which describes localized states in a cholesteric (the coefficients A, B, and C are given by Eqs. (29), and R has the significance of a characteristic radius of the filament core).

The functional dependences of the potentials (28) and (29) correspond to the transformation properties of the system energy (15) discussed in the preceding section. The coefficient A includes the contribution of the regular (nematic) part of the elastic energy as described by terms quadratic in the spatial derivative. This energy is always positive and makes the energy of the inhomogeneous state higher than that of the homogeneous state. In contrast, the invariants linear in the spatial derivatives, being attributable to the chirality of the system, provide a negative contribution to the energy of modulated states. In the model energy (28) their contribution is represented by the term linear in R, which tends to stabilize the system for finite lattice dimensions.

The balance of the indicated opposing contributions to the system energy determines the equilibrium parameters of the inhomogeneous structures. The inhomogeneity of the system increases upon transition from one-dimensional modulated structures to two dimensions (lattices). The absolute values of the competing contributions to the system energy increase in this case: on the one hand, the value of the coefficient B increases and, on the other, the ''stiffness'' of the system (the value of the parameter A) increases. As shown above, the equilibrium energy of a helicoid is lower than the energy of lattices. Consequently, in contrast to inhomogeneous states in the mixed state of superconductors, where a two-dimensional modulated structure (lattice of Abrikosov vortices) is stable,[31,32] in cholesterics the system of minimum energy is one with moderate inhomogeneity, i.e., a one-dimensionally modulated structure (helicoid).

It follows from an analysis of the potential (28) that the equilibrium radii of π and $\pi/2$ lattices (R_0) do not depend on the field and, accordingly, are $R_0^{(1)}=2A_1/B_1=(1.079+0.168k)$ and $R_0^{(2)}=2A_2/B_2=(0.732+0.150k)$. Outside the immediate vicinity of the critical fields the distribution $\theta(r)$ in the cell (Fig. 9) exhibit more or less regular behavior. The results obtained above in the ''linear'' approximation give a correct qualitative description of the evolution of the system in this region (in particular, the lattice dimensions increase as k increases, and they are independent of the magnetic field away from the critical fields); moreover, there is reasonably good quantitative agreement with the results of numerical calculations.

It is significant that the dimensions of localized structure do not depend on A, but the ''stiffness'' of the system is governed by the energy of interaction with the field. The equilibrium radius of the core of a filament $R_0=B_1/(2C_1)\sim 1/h^2$ increases without bound in the limit $h\to 0$ (cf. Fig. 7). A potential of the form (30) has been obtained in Ref. 35, in which the first study of axial structures in systems without an inversion center is reported.

Consequently, the model potentials (28) and (30) can be used to analyze the physical causes responsible for the stabilization of axial structures; the majority of the results obtained in the given approximation are confirmed by a more rigorous treatment and fall within a fairly broad class of field models (see Ref. 34).

For $k=1$ the energy (15) functionally coincides with the energies of axial structures (magnetic vortices) in an easy-axis ferromagnet without an inversion center in zero field.[27–29] The energy of a magnetic vortex in a uniaxial antiferromagnet has the same functional form.[36] As in cholesterics, in antiferromagnets without an inversion center in the modulated state a helicoid structure is always thermodynamically stable, and vortex lattices can exist as metastable formations.[36]

We conclude this section with the observation that topological and structural analogs of inhomogeneous configurations of $\mathbf{n}(\mathbf{r})$ in liquid crystals can be found in various nonlinear field models of the physics of the condensed state and the theory of elementary particles. For example, the distribution of the magnetization in a helicoid structure of an anisotropic magnet without an inversion center[22] is the same as the distribution of the director in a cholesteric in an external field.[21] The Kosterlitz–Thouless vortex structures in an easy-

plane Heisenberg ferromagnet[37] are analogous to planar disclinations of index 2 in nematics. The structural and topological similarity between inhomogeneous configurations in antiferromagnets and nematics is well known (see, e.g., Ref. 13). The two-dimensional axial structures with a nonsingular core discussed in the present paper are analogous to vortex states in superconductors,[31,32] in liquid He^4 (Ref. 38), and in magnets.[28,29] An analogy exists between three-dimensional defects in liquid crystals (drops) and monopoles in the theory of gauge fields.[39,40]

The stated similarity in the basic properties of localized states (solitons) and nonlocalized states in various physical models, on the one hand, is based on the common topological properties of such structures[41] and, on the other, is a consequence of general theorems on the character of the solutions for various types of nonlinear differential equations.[34,42]

A liquid crystal has several advantages over other physical systems for the modeling and investigation of various inhomogeneous structures. The system parameters can be varied over wide limits to establish the necessary conditions for a given experiment; as a rule, the experiments are conducted at room temperature and are comparatively simple; the results of investigations are easily visualized, to a degree not usually attainable in the investigation of other nonlinear media.

The results of the theoretical investigation and some experimental data lead to the assumption that cholesterics in an external field model a uniaxial ferromagnet or antiferromagnet without an inversion center, while a nematic models an isotropic ferromagnet or antiferromagnet. Moreover, in this paper (see also Refs. 29 and 34) we have shown that certain properties of the inhomogeneous states of the investigated inhomogeneous states of liquid crystals are not associated with a specific model, but are characteristic of a broad class of systems and therefore have a universal character. Consequently, the results obtained in the investigation of axial structures in liquid crystals should be useful in other branches of physics of the condensed state (primarily in magnetism) and in the theory of solitons.

4. POSSIBLE EXPERIMENTS AND APPLICATIONS

The applied physics of liquid crystals has begun to develop at a vigorous pace through the application of electrooptical field effects in optical devices using nematics.[43] Despite impressive progress, the subsequent refinement of liquid-crystal devices (primarily color displays) will require the solution of a host of complex technological, engineering, and fundamental problems.[44]

Research aimed specifically at axial structures in liquid crystals has not been carried out to date. We have cited studies of nematic structures in cylindrical capillaries[10,11] and a number of papers reporting the observation of two-dimensional modulated structures in cholesterics in a definite range of electric and magnetic fields.[14–18] Two-dimensional lattices induced by stresses and temperature gradients have also been observed in cholesterics.[15] Two-dimensional modulated structure are also induced in chiral smectics under the influence of stresses and light beams (Kahn effect).[1,45] Spatially periodic structures of this kind have been treated theoretically as slight periodic distortions of a helicoid (Helfrich–Hurault effect).[46,47]

According to the results of the present study, static two-dimensional modulated states in the form of lattices of axial structures are possible in liquid crystals without an inversion center. These states are separated from the helicoid phase by a finite potential barrier and are impossible to obtain by the continuous deformation of a helical structure. Unfortunately, so far no one has established the distribution $\mathbf{n}(\mathbf{r})$ in the bulk of materials in which two-dimensional modulated structures are observed. It would be important to conduct such investigations and also to analyze the influence of external fields on the lattice parameters. According to the results of the present study, weak external fields can induce appreciable regular distortions of the field of the director $\mathbf{n}(\mathbf{r})$ in the lattice. Research of this kind could therefore have practical applications.

Since the helicoid phase is thermodynamically stable in the entire region of existence of modulated structures, special conditions are necessary for the induction of lattices. For example, in Refs. 14–18 the dimensions of the planar unit cell and the character of the interaction of a cholesteric with the support surfaces have played an important role in the formation of two-dimensional structures. Consequently, one direction in which to expand the investigations reported here entails the calculation of modulated structures in plane capillaries of finite thickness, taking into account the influence of horizontal support surfaces.

Also promising from the standpoint of applications are investigations of cholesteric axial structures in cylindrical capillaries. As shown in the present study, the distribution $\mathbf{n}(\mathbf{r})$ in cholesterics is easily altered by external fields, a fact that can exploited in various optical devices. We recall that nematic axial structures with a continuous distribution $\mathbf{n}(\mathbf{r})$ have been observed experimentally in cylindrical capillaries.[10,11]

We have discussed the relationship between the liquid-crystal axial structures discussed here and vortex states of magnets without an inversion center. In a ferromagnet in a magnetic field the energy includes a term linear in the magnetization. Interaction with an external field reduces the energy equivalence of antiparallel directions of the structure vector (the magnetization in our case). This effects affords an added possibility of regulating the structure of modulated states. The evolution of ferromagnetic vortex lattices and solitary vortices in a magnetic field has been studied in several papers.[27–29,35] In particular, it has been determined that vortex lattices can be thermodynamically stabilized in a certain range of fields. A lucrative undertaking in this regard is the investigation of axial structures in chiral liquid crystals exhibiting spontaneous polarization (e.g., in smectics S_c^*; see Refs. 3, 44, and 48) and in suspensions of ferroparticles in cholesterics. Judging from the results of Refs. 27–29 and 35 and data on the properties of ferroelectric liquid crystals[48] and ferronematics,[49] in such systems we can anticipate the

formation of thermodynamically stable lattices of axial structures, which are easily controlled by external fields, in such systems.

5. CONCLUSIONS

1. The equations of the phenomenological theory of elasticity for nematics have solutions in the form of axisymmetric localized states with a continuous distribution of the director $\mathbf{n}(\mathbf{r})$ (6); axial structures with $\mathbf{n}$ rotated through an angle less than 180° can exist in cylindrical capillaries of nematics. These solutions become radially unstable in external fields. In a magnetic field the axial structures are compressed into dense filaments, in which the distances between molecules are comparable with their diameters. The collapse of axial structures also takes place in nematics with positive dielectric anisotropy; however, if $\Delta\varepsilon<0$, radial instability is manifested in the unbounded expansion of the given axial formations. In our ideal model the instability has a nonthreshold character and will necessarily occur in an indefinitely small field. According to experimental data, real systems must be characterized by a threshold field, below which an axial structure is stable.

2. ''Dark filaments,'' the most common type of nematic texture, are possibly localized axial structures compressed by random external fields. Evidence in support of this model is the fact that axial structures, unlike disclinations, have finite energy. It is also known that free dark filaments have a high mobility; the motion of disclinations, on the other hand, is associated with energy expenditures due to discontinuities in the field of the director $\mathbf{n}(\mathbf{r})$.

3. Coexisting with one-dimensional modulated states (helicoids) in cholesterics are lattices of axial structures. Two types of lattices are possible according to whether the director is rotated through 90° or 180°. The equilibrium dimensions of the unit cells are comparable to the period of the helicoid. The lattices have finite dimensions throughout the entire region in which the modulated state exists; their energies are lower than the energy of the homogeneous state, but are slightly higher than the energy of the helicoid phase. All modulated states (helicoid and both lattice type) go over to the homogeneous state through the unbounded growth of the period in a critical magnetic field H_c (13).

4. A system of localized structures can exist in cholesterics above the critical field H_c. As the field is increased, the effective dimensions of such formations decrease, but even in very strong fields they do not collapse.

5. Axial structures with a core of finite radius can exist in cylindrical capillaries of cholesterics. Even low external fields induce substantial regular variations of the distribution $\mathbf{n}(\mathbf{r})$ in such structures.

6. The new types of inhomogeneous states obtained in the present study in liquid crystals readily change structure and, hence, optical properties under the influence of a field. A benefit of this behavior is the potential for practical applications of both isolated axial structures and lattices of such structures.

Localized axial structures and lattices of such structures can also exist in chiral liquid crystals exhibiting spontaneous polarization (e.g., in ferroelectric liquid crystals). Lattices of axial structures can become thermodynamically stable in such systems in a certain range of fields.

The authors are grateful to the Alexander von Humboldt Foundation for furnishing computer facilities and literature on liquid crystals. We are deeply indebted to V. G. Bar'yakhtar and A. Hubert for a discussion of this work and for valuable consultations, and to A. S. Kovalev and S. V. Tarasenko for useful discussions.

*)E-mail: bogdanov@host.dipt.donetsk.ua

[1] P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, 2nd ed., Clarendon Press, Oxford (1993).
[2] W. H. de Jen, *Physical Properties of Liquid Crystalline Materials*, Gordon and Breach, New York (1980).
[3] *Liquid Crystals*, H. Stegemeyer (Ed.), Steinkopff Verlag, Darmstadt; Springer, New York (1994).
[4] I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **58**, 1443 (1970) [Sov. Phys. JETP **31**, 773 (1970)].
[5] M. J. Stephen and J. P. Straley, Rev. Mod. Phys. **46**, 618 (1974).
[6] A. S. Pikin, *Structural Transformations in Liquid Crystals* [in Russian], Nauka, Moscow (1981).
[7] S. Chandrasekhar, *Liquid Crystals*, Cambridge Univ. Press, New York (1977).
[8] *Liquid Crystals (Applications and Uses)*, Vols. 1–3, (Ed.), B. Bahadur World Scientific, Singapore–River Edge, N.J.–London (1990).
[9] P. E. Cladis and M. Kleman, J. Phys. (Paris), Colloq. **33**, 591 (1972).
[10] R. B. Meyer, Philos. Mag. **27**, 403 (1973).
[11] C. Williams, P. Pieranski, and P. E. Cladis, Phys. Rev. Lett. **29**, 90 (1973).
[12] S. I. Anisimov and I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **61**, 2140 (1971) [Sov. Phys. JETP **34**, 1140 (1972)].
[13] C. Rebbi and G. Soliani, *Solitons and Particles*, World Scientific, Singapore (1984).
[14] F. Rondelez and J. P. Hulin, Solid State Commun. **10**, 1009 (1972).
[15] H. Hervet, J. P. Hurault, and F. Rondelez, Phys. Rev. A **8**, 3055 (1973).
[16] F. Rondelez and H. Arnould, C.R. Acad. Sci. Ser. B **273**, 549 (1971).
[17] G. Malet, J. Marignan, and O. Parodi, J. Phys. Lett. **36**, 317 (1975).
[18] J. Marignan, G. Malet, and O. Parodi, J. Phys. Lett. **37**, 3365 (1975).
[19] A. N. Bogdanov and A. A. Shestakova, Fiz. Nizk. Temp. **23**, 1024 (1997) [Low Temp. Phys. **23**, 769 (1997)].
[20] A. A. Belavin and A. M. Polyakov, JETP Lett. **22**, 245 (1975).
[21] P. G. De Gennes, Solid State Commun. **6**, 163 (1968).
[22] I. E. Dzyaloshinskiĭ, Zh Éksp. Teor. Fiz. **46**, 1420 (1964) [Sov. Phys. JETP **19**, 960 (1964)].
[23] E. Sackmann, S. Melboom, and L. C. Snyder, J. Am. Chem. Soc. **89**, 5982 (1967); J. Wysocki, J. Adams, and W. Haas, Phys. Rev. Lett. **20**, 1025 (1968).
[24] G. Durand, L. Leger, F. Rondelez, and M. Veyssle, Phys. Rev. Lett. **22**, 227 (1969).
[25] R. B. Meyer, Appl. Phys. Lett. **14**, 208 (1969).
[26] F. J. Kahn, Phys. Rev. Lett. **24**, 209 (1970).
[27] A. N. Bogdanov, M. V. Kudinov, and D. A. Yablonskiĭ, Fiz. Tverd. Tela (Leningrad) **31**(10), 99 (1989) [Sov. Phys. Solid State **31**, 1707 (1989)].
[28] A. N. Bogdanov and A. Hubert, J. Magn. Magn. Mater. **138**, 255 (1994).
[29] A. N. Bogdanov and A. Hubert, Phys. Status Solidi B **186**, 527 (1994).
[30] A. Hubert, *Theorie der Domanenwande in Geordneten Medien* Springer-Verlag, Berlin (1974).
[31] A. A. Abrikosov, Zh. Éksp. Teor. Fiz. **32**, 1442 (1957) [Sov. Phys. JETP **5**, 1174 (1957)].
[32] P. G. De Gennes, *Superconductivity of Metals and Alloys*, Benjamin, New York (1966).
[33] V. A. Belyakov and V. E. Dmitrienko, Usp. Fiz. Nauk **146**, 369 (1985) [Sov. Phys. Usp. **28**, 535 (1985)].
[34] A. N. Bogdanov, JETP Lett. **62**, 247 (1995).
[35] A. N. Bogdanov and D. A. Yablonskiĭ, Zh. Éksp. Teor. Fiz. **95**, 178 (1989) [Sov. Phys. JETP **68**, 101 (1989)].
[36] A. Bogdanov, in *Sixth Joint MMM-Intermag Conference, Abstracts*, Albuquerque (1994).
[37] J. M. Kosterlitz and D. J. Thouless, J. Phys. C Solid State Phys. **6**, 1181 (1973).

[38] R. J. Donnelly, *Quantized Vortices in Helium II*, Cambridge Univ. Press, Cambridge–New York (1991).
[39] M. V. Kulik and O. D. Lavrentovich, Zh. Éksp. Teor. Fiz. **85**, 511 (1983) [Sov. Phys. JETP **58**, 299 (1983)].
[40] O. D. Lavrentovich and E. M. Terent'ev, Zh. Éksp. Teor. Fiz. **91**, 2084 (1986) [Sov. Phys. JETP **64**, 1237 (1986)].
[41] G. E. Volovik and V. P. Mineev, Zh. Éksp. Teor. Fiz. **72**, 2256 (1977) [Sov. Phys. JETP **45**, 1186 (1977)].
[42] V. G. Makhan'kov, Yu. P. Rybakov, and V. I. Sanyuk, Usp. Fiz. Nauk **164**/, 121 (1994).
[43] M. Schadt and W. Helfrich, Appl. Phys. Lett. **18**, 127 (1971).
[44] M. Schadt, in *Liquid Crystal*, H. Stegemeyer (Ed.), Steinkopff Verlag, Darmstadt; Springer, New York (1994).
[45] F. Kahn, Appl. Phys. Lett. **22**, 111 (1973).
[46] W. Helfrich, Appl. Phys. Lett. **17**, 531 (1970).
[47] J. P. Hurault, J. Chem. Phys. **59**, 2068 (1973).
[48] L. M. Blinov and L. A. Beresnev, Usp. Fiz. Nauk **143**, 391 (1984) [Sov. Phys. Usp. **27**, 492 (1984)].
[49] Yu. L. Raĭkher, S. V. Burylov, and A. N. Zakhlevnykh, Zh. Éksp. Teor. Fiz. **91**, 542 (1986) [Sov. Phys. JETP **64**, 319 (1986)].

Translated by James S. Wood

Magnetic properties and energy transfer in the luminophors CaS:Eu,Cl

S. O. Klimonskiĭ, A. É. Primenko, and V. D. Kuznetsov

D.I. Mendeleev Russian Chemical Engineering University, 125820 Moscow, Russia

M. I. Danilkin and V. Seman

University of Tartu, 202400 Tartu, Estonia
(Submitted 19 May 1997)
Zh. Éksp. Teor. Fiz. **113**, 1698–1707 (May 1998)

Studies have been performed on the magnetic properties and electron paramagnetic resonance (EPR) spectra of the small-crystal luminophors CaS:Eu,Cl. By comparing the EPR and magnetic susceptibility data we have determined that Eu enters into the CaS matrix primarily in the trivalent state Eu^{3+}. We have found that the magnetic susceptibility of the sample with the lowest europium content (5×10^{-3} at. %) has a sharp peak at $T\approx5$ K. A model is proposed of clusters into which the europium ions in these luminophors can associate. On the basis of this model an explanation is given for the anomalous temperature dependence of the magnetic susceptibility as well as a long list of other experimental facts (including peculiarities of the thermal luminescence). © *1998 American Institute of Physics.* [S1063-7761(98)01005-1]

1. INTRODUCTION

Studies of luminescence and the associated energy transfer are a vigorously developing branch of research. The determination of the mechanisms of these processes is of fundamental scientific interest for solid state physics. One object of such research is europium activated CaS, which has a characteristic luminescence band corresponding to $5d-4f$ transitions in Eu^{2+}. To elucidate the luminescence mechanism in CaS:Eu it is necessary to elucidate the nature of the radiating centers and the path of energy transfer to them. The task is complicated by the fact that the energy transfer mechanism depends on the number of intrinsic defects, additional impurities and complexes formed of defects and impurity atoms. Thus, for the luminophor CaS:Eu,Sm the following mechanism of thermal luminescence is proposed:[1]

—liberation of electrons from samarium ions (playing the role of electron traps) $Sm^{2+}=Sm^{3+}+e^-$,

—recapture by Eu^{3+}: $Eu^{3+}+e^-=Eu^{2+}$,

—luminescence of the excited Eu^{2+} ions.

In the case when the additional impurity is chlorine (CaS:Eu,Cl), the main thermal luminescence peak turns out to be associated with liberation of holes rather than electrons. In this case the fraction of europium ions found in the state Eu^{2+} is almost independent of illumination of the material.

For an isolated europium ion replacing a Ca^{2+} ion, the Eu^{2+} charge state is preferred. However, in this case strong lattice distortions arise since the Eu^{2+} ion is substantially larger than Ca^{2+} (the ionic radii are respectively 1.17 and 1.00 Å). Therefore the alternative state Eu^{3+} is possible (ionic radius 0.95 Å) if other impurities and defects uphold the principle of charge compensation. The europium distribution over the two charge states depends on the conditions of synthesis of the luminophor. It is necessary to know this distribution in order to understand the structure of the europium centers. Magnetic methods can be very helpful here since the magnetic properties of the Eu^{2+} and Eu^{3+} ions differ sharply: Eu^{3+} is found at low temperatures ($T<200$ K) in the nonmagnetic state $^7F_0$ and is characterized by very weak temperature-independent van Vleck paramagnetism. Even at room temperature it gives hardly any contribution to the electron paramagnetic resonance (EPR) and magnetic susceptibility. At the same time, Eu^{2+} is found in the $^8S_{7/2}$ state with purely spin magnetic moment $S=7/2$ and has a characteristic EPR center and magnetic susceptibility obeying a Curie law.

2. EXPERIMENTAL METHODS

In the present work we combined magnetic susceptibility and EPR studies of powdered small-crystal samples of CaS:Eu,Cl with varying europium ($0.005<x<0.16$ at. %) and chlorine content ($0<y<0.32$ at. %). To prepare them, we first synthesized CaS by reducing $CaSO_4$ in a hydrogen flux at 900 °C, where the initial $CaSO_4$ was prepared by precipitation from solutions of $(NH_4)_2SO_4$ and $Ca(NO_3)_2$ first washed with organic solvents to remove transition metal impurities.[2] The synthesized high-purity CaS was then mixed with the activating impurities Eu_2O_3 and $CaCl_2$ and annealed in sulfur vapors at 1100 °C. As a result, we obtained samples containing around $0.5-1$ at. % O, $\approx10^{-3}-10^{-2}$ at. % residual impurities Cl, F, Na, Li and less than 10^{-6} at. % Cr, Co, Fe, Pb, and some other elements. All samples had a thermally stimulated luminescence spectrum similar to that described earlier,[2] whose main peak (at 210 K) is associated with liberation of holes.

The EPR measurements of the CaS:Eu,Cl samples were performed at room temperature with the help of an RÉ-1301 EPR spectrometer at 9.1 GHz. The magnetic susceptibility measurements were performed on a SQUID magnetometer[3] in a constant magnetic field $H=777$ Oe at temperatures $2\text{ K}\leqslant T\leqslant50$ K. In preparation for the measurements, the

TABLE I.

Sample No.	Batch concentration		Curie constant C, 10^{-6}, K/g	Eu^{2+} ion concentration	
	χ (Eu), at.%	y(Cl), at.%		x^{2+}, at.%	x^{2+}/x
1	0.005	–	2.6 ± 0.1	0.0024	0.47
2	0.010	–	3.1 ± 0.1	0.0029	0.29
3	0.020	–	6.5 ± 0.1	0.0060	0.30
4	0.040	–	13.4 ± 0.1	0.0123	0.31
5	0.080	–	22.5 ± 0.1	0.0206	0.26
5/16	0.080	0.16	25.4 ± 0.1	0.0232	0.29
5/24	0.080	0.24	39.7 ± 0.1	0.0363	0.45
5/32	0.080	0.32	41.9 ± 0.3	0.0384	0.48
6	0.160	–	39.1 ± 0.2	0.0358	0.22

samples were pumped down and sealed in quartz ampoules with a thin partition making a negligibly small and practically temperature-independent contribution to the reading. The gaseous medium needed for rapid establishment of thermal equilibrium was created by diffusing helium through the walls of the ampoules. For a mass of the charges around 100 mg the error of the magnetic susceptibility χ measurements was on the order of $\pm 5\times 10^{-9}$ g^{-1}.

3. EXPERIMENTAL RESULTS AND DISCUSSION

We investigated the samples indicated in Table I of CaS:Eu in which special care was taken not to introduce chlorine, samples additionally doped with chlorine, and a control sample containing neither Eu nor Cl. The susceptibility of the last sample had quite a weak temperature dependence, with $\chi_{CaS} \approx -3\times 10^{-7}$.

For most of the samples not containing Cl (i.e., containing Cl only at the level of a residual impurity), besides samples No. 1 and 2 with the highest Eu content ($x=0.01$ and 0.005 at. %), the magnetic susceptibility data were well rectified in coordinates $\chi - T^{-1}$ (Fig. 1), which corresponds to the dependence $\chi(T)=\chi_0+C/T$, where the first (temperature-independent) term characterizes the susceptibility of the crystalline matrix ($\chi_0 \approx \chi_{CaS}$), and the second is the nearly Curie-law contribution of the Eu^{2+} ions. The corresponding values of C, determined for $T>7$ K, are given in Table I, which, along with Fig. 2, give the values of x^{2+}, the concentration of europium in the ''2+'' state, calculated from C. In the chlorine-free samples the fraction of europium in this state falls with growth of its total concentration. This may be due to a growth of strains of the crystal lattice resulting from the large ionic radius of Eu^{2+}. At large x values, up to 80% of the europium turns out to be in the Eu^{3+} state and is found in the crystal primarily in the charge-compensated complexes $V_{Ca}^{-2}Eu^{3+}$ (where V denotes a vacancy). Chlorine doping leads to an increase in the Eu^{2+} fraction, which is entirely natural since the substitution $S^{2-}\rightarrow Cl^{-}$ leaves an excess electron. It is interesting that the influence of small Cl concentrations on the Eu^{2+} content is quite weak; however, already at 0.24 at. % Cl the Eu^{2+} concentration grows by almost a factor of two. This result confirms the conclusion drawn in Ref. 4 that at small concentrations ($y<0.2$ at. %) Cl enters into CaS mainly in the form of $[V_{Ca}Cl_S^-]^{(-)}$ acceptor centers, and only later do isolated $[Cl_S^-]^{(+)}$ ions appear.

For samples No. 2 ($x=0.01$ at. %) and No. 5/24 ($x=0.08$ at. %, $y=0.24$ at. %) at temperatures below 5 K deviations (toward a decrease) from a Curie law were observed. This indicates the presence of an anomalous component, whose contribution, however, is barely noticeable against the background of the susceptibility of the Eu^{2+} ions, which rises rapidly as the temperature is lowered. This effect is most distinctly observed in sample No. 1 ($x=0.005$ at. %).

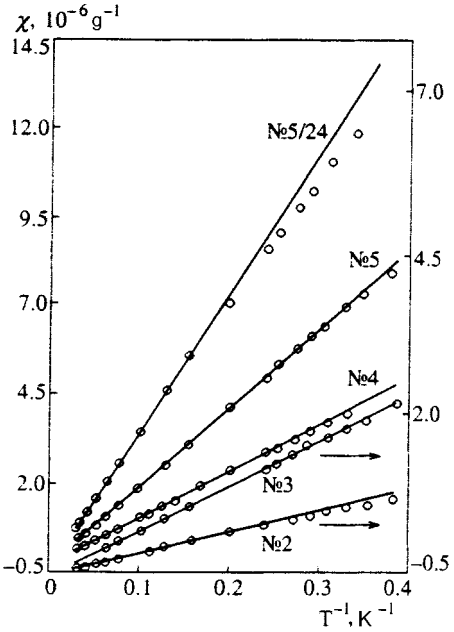

FIG. 1. $\chi(T^{-1})$ curves: No. 2, 3—right axis, No. 4, 5 and 5/24—left axis. Curve numbers correspond to sample numbers in Table I. The straight lines approximate the experimental data for $T>7$ K.

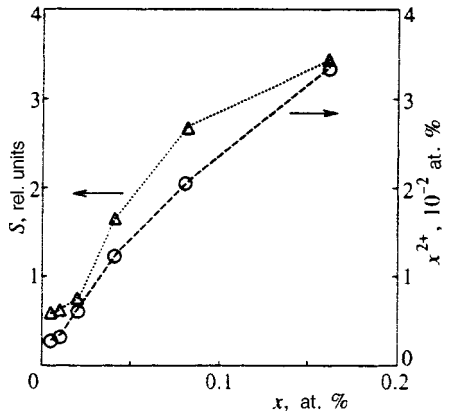

FIG. 2. Dependence of the area S under the Eu^{2+} EPR spectrum and the Eu^{2+} ion concentration (calculated from the Curie constant C) on the incorporated europium concentration for non-chlorine-doped samples.

For this sample the dependence $\chi(T)$ clearly deviates from a Curie law already at $T<15$ K, and at $T\approx 5$ K the paramagnetic susceptibility passes through a maximum, falling abruptly with further reduction of the temperature (Fig. 3). Such an anomaly can have the following explanations:

1) the disappearance of the paramagnetic susceptibility of Eu^{2+} below 5 K is the result of an antiferromagnetic interaction (possibly an antiferromagnetic phase transition) in clusters or inclusions that contain europium;

2) almost all of the europium is found in the state Eu^{3+}, and peculiarities in the $\chi(T)$ curves are due to other paramagnetic centers associated with residual impurities and intrinsic defects;

3) at high temperatures some of the europium is found in the state Eu^{2+}; however, at temperatures below 5 K it loses an electron to a neighboring defect and transitions to the nonmagnetic state Eu^{3+}.

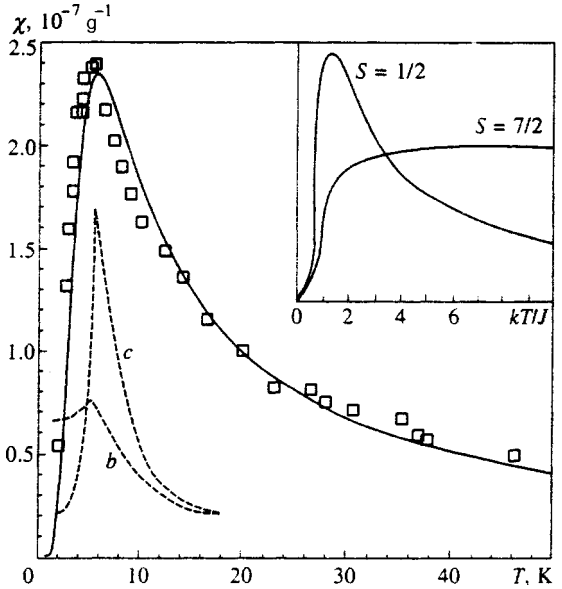

FIG. 3. Temperature dependence of the magnetic susceptibility for sample No. 1. The solid curve was calculated for $\Delta E\approx 13$ K (see text). For comparison, the dependence $\chi(T)$ is shown for the Eu_3O_4 phase in the direction of the b and c axes.[9] Inset: calculated curves for antiferromagnetically bonded pairs for the spins $S=1/2$ and $S=7/2$ (in arbitrary units).

Let us consider these three possibilities in sequence.

In case (1) the most likely candidate is the phase Eu_3O_4, which is an antiferromagnet[5] with Néel temperature $T_N\approx 5$ K. This hypothesis is supported by the appreciable content ($\approx 0.5-1\%$) of residual oxygen in the samples (see above). However, x-ray diffraction studies carried out by us on similar samples with somewhat greater europium concentration ($x>0.2$ at. %) showed that the the nonmagnetic phase that is observed to precipitate out is Eu_2O_2S, not Eu_3O_4. Furthermore, the decrease of the susceptibility of this compound below the Néel temperature is extremely insignificant when the magnetic field is aligned with the hard magnetization axis.[6] After averaging over all crystallographic directions in a powdered sample, the depth of falloff of $\chi(T)$ should not exceed 50%. In the temperature range $4<T<6$ K, just where a sharp peak in $\chi(T)$ should be observed for Eu_3O_4 (Fig. 3), we instead observe a wide maximum, and as the temperature is lowered to 2 K the susceptibility decreases by roughly fivefold, i.e., significantly more strongly than even in ideal samples of the phase Eu_3O_4.

Speaking of antiferromagnetic interactions, we should also recall the possible existence of Eu^{2+} antiferromagnetic clusters. Among such clusters only antiferromagnetic pairs can give a distinct maximum in the $\chi(T)$ dependence.[7] However, to use them to explain the anomalous behavior observed for sample No. 1 is also impossible. The reasons for this are the following. First, for the Eu^{2+} susceptibility to fall abruptly at $T\approx 2$ K the exchange interaction parameter in the pairs should have the value $J/k_B\approx -4$ K, which is an order of magnitude greater than the values characteristic of europium chalcogenides (in EuS $J/k_B=0.2$ and -0.08 K respectively for the first and second coordination spheres[5]). Second, the greater the spin of the centers bound to the antiferromagnetic pairs, the wider and flatter the maximum in $\chi(T)$ (Ref. 7). The dependence actually observed for sample No. 1 is similar in shape to the curve corresponding to pairs of particles with spin $S=1/2$, but corresponds poorly to the curve for pairs of Eu^{2+} ions with $S=7/2$ (see the inset to Fig. 3). Thus, if the anomalous dependence of $\chi(T)$ can be connected with antiferromagnetic pairs, then those pairs can only be Eu^{2+} pairs.

Let us now consider case (2), which assumes that all of the europium is found in the Eu^{3+} state. To test this hypothesis it is necessary to enlist additional experimental data, e.g., EPR spectra.

EPR spectra of all the chlorine-free samples are very similar (see Fig. 4): they all contain sharp, narrow lines corresponding to the hyperfine structure of the two isotopes $^{151}Eu^{2+}$ and $^{153}Eu^{2+}$ located at positions with nearly octahedral local symmetry, and also a wide, smooth line superimposed on them. This line results primarily from broadening and overlap of lines from Eu^{2+} ions located in a non-octahedral environment upon averaging of the signal over grains with different orientations.[4]

To compare the Eu^{2+} concentration in the various samples we calculated the total area S under all the EPR lines for each sample. The results are plotted in Fig. 2. At

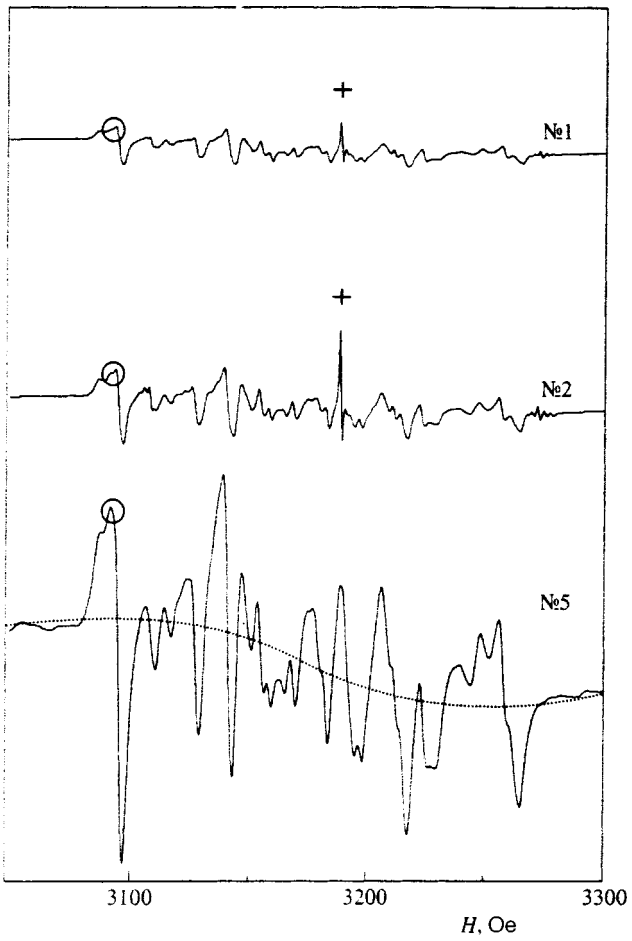

FIG. 4. EPR spectra for samples No. 1, 2, and 5. The broad line, hidden under the narrow peaks, is shown for sample No. 5.

large x the dependence $S(x)$ is sublinear, in agreement with the magnetic susceptibility data, and indicates a decrease of the Eu^{2+} fraction with growth of x. Note that the nonlinearity of $S(x)$ is evident even at small x. Such behavior can be linked with the influence of residual impurities on the charge state of the europium ions. Thus we see that the hypothesis of the absence of Eu^{2+} in sample No. 1 is not confirmed; moreover, the EPR spectra show that the Eu^{2+} fraction in this sample is higher than in the others. Additional confirmation of the presence of Eu^{2+} is provided by the presence of thermal luminescence in all the samples (see above).

It follows from the above that the anomalous behavior of the magnetic susceptibility for sample No. 1 should be linked with the Eu^{2+} ions because there is no other paramagnetic contribution that could be linked with europium in this sample. Moreover, if we were dealing with paramagnetism of any other centers, they would have been detected in the sample not containing europium, which did not happen.

Hypothesis (3), which we consider the most likely, is that in sample No. 1 below 5 K one of the electrons jumps from the Eu^{2+} ion to a neighboring defect. As a result, the magnetic Eu^{2+} ion is converted into a Eu^{3+} ion with zero magnetic moment. In the other samples having anomalies in the dependence below 5 K a similar conversion is observed, obviously, for only some of the Eu^{2+} ions.

What kind of defect adjacent to the Eu ion are we dealing with? First note that the formation of Eu-containing complexes with anomalous magnetic susceptibility is enhanced in the chlorine-doped samples; consequently, Cl should be part of such complexes. As residual impurities, Cl and F in quantities up to $10^{-3}-10^{-2}$at. % are also present in the intentionally undoped samples, and this quantity can be enough to completely bond the europium in sample No. 1. Second, at low concentrations Cl enters into CaS in the form $V_{Ca}Cl_S^-$ centers,[4] and consequently we can conclude with confidence that Eu is present in complexes with $V_{Ca}Cl_S^-$ or $V_{Ca}F_S^-$ centers. This is indicated by the fact that for sample No. 5/24 ($x=0.080$ at. %, $y=0.24$ at. %) a more pronounced deviation from the Curie law was observed than for sample No. 5/32 ($x=0.080$ at. %, $y=0.32$ at. %) since the largest number of $V_{Ca}Cl_S^-$ centers is formed precisely at $y=0.2$ at. %.

As a criterion of the validity of the hypothesis that complexes are formed with recharging europium ions we cite the conformity of such a hypothesis with the properties of sample No. 1, since there are no other paramagnetic centers in this sample (in the remaining samples isolated Eu^{2+} ions are probably also present). As a candidate for such a complex we propose the following cluster:

$$Eu_{Ca}^{3+}-V_{Ca}^{(2-)}-Cl_S^-Eu_{Ca}^{3+} \quad (1)$$

(we have indicated here its low-temperature ground-state charge configuration, which does not have a magnetic moment). It differs from the complex $Eu^{3+}-V_{Ca}^{(2-)}-Eu^{3+}$, in which Eu^{3+} ions can be found,[8] by the fact that in it Cl^- ions substitute at sulfur sites. As a result, the complex is positively charged relative to the lattice and requires charge compensation, which can be provided by the residual impurity ions $[Na_{Ca}^+]^-$, $[Li_{Ca}^+]^-$, and $V_{Ca}^{(2-)}Cl_S^-$ centers and other acceptors.

We believe that above 2 K the given complex can have another charge configuration:

$$Eu_{Ca}^{3+}-V_{Ca}^{(-)}-Cl_S^-Eu_{Ca}^{2+}. \quad (2)$$

The configuration has a large magnetic moment associated with the Eu^{2+} ion and the hole residing at the vacancy $V_{Ca}^{(-)}$. At high temperatures the complex should easily transition from the one configuration into the other. This is facilitated by the Cl^- ions, which are positively charged relative to the lattice and lower the height of the barrier the electron has to overcome to make the transition. The probability of the realization of each of configurations (1) and (2) should be determined by their statistical weight and the Boltzmann factor $\exp(-\Delta E/k_BT)$, where ΔE is the energy difference between the two configurations. When all of the europium is bound up in complexes (1) and (2), as is probably the case for sample No. 1, it is not hard to calculate the corresponding temperature dependence of the magnetic susceptibility:

$$\chi_p=\frac{C}{T}\left(1+\frac{n_1}{n_2}\exp\left[-\frac{\Delta E}{k_BT}\right]\right)^{-1},$$

$$C=\frac{\mu_B N_A x}{k_B M_{CaS}}\left(g_{Eu}^2 S_{Eu}(S_{Eu}+1)+g_h^2 S_h(S_h+1)\right),$$

Here n_1 and n_2 are the statistical weights of complexes (1) and (2), respectively, k_B is the Boltzmann constant, μ_B is the Bohr magneton, N_A is Avogadro's number, x is the atomic fraction of europium, M_{CaS} is the molecular weight of CaS, g_{Eu} and g_h are the g factors of the Eu^{2+} ion and the hole, respectively, and $S_{Eu}=7/2$ and $S_h=1/2$ are their respective spins. In the calculation we neglected the interaction between the europium ion and the hole, and we took the two g factors g_{Eu} and g_h to be equal to 2. The statistical weight for complex (1) $n_1=1$, and for complex (2) $n_2=16$ (the eightfold degenerate europium ion complexed with a doubly degenerate hole). The quantity ΔE was used as a fitting parameter. As a result of fitting the experimental data, we obtained the result $\Delta E \approx 13$ K (see Fig. 3). It can be seen from this figure that the agreement between the calculated curve and the experimental points is entirely satisfactory.

According to the calculation, the susceptibility at high temperatures is governed by configuration (2). Taking its large statistical weight into account, it turns out that at high temperatures 47% of the Eu ions should be found in the 2 + state. This number is found to be in good agreement with the experimental data (see Table I). At low temperatures, on the contrary, the complexes transition back to the ground state (1), and their susceptibility essentially vanishes. In the intermediate temperature region $T>10$ K the dependence can be formally approximated by a Curie–Weiss law $\chi-\chi_{CaS}=C/(T-\theta)$ with negative temperature parameter $\theta\approx -6$ K, which within the context of the proposed model characterizes not an antiferromagnetic interaction, but a decrease in the paramagnetism as the temperature is lowered as a consequence of the transition to the nonmagnetic state.

As has already been noted, sample No. 1 exhibits thermally stimulated luminescence. Therefore, let us consider how the proposed complexes can store light energy. This can happen as a result of the ejection of one electron in configuration (1) from a vacancy to Eu^{3+}:

$$Eu_{Ca}^{3+}-V_{Ca}^{(2-)}-Cl_S^-+Eu_{Ca}^{(3+)}+h\nu \rightarrow Eu_{Ca}^{2+}-V_{Ca}^{(-)}-Cl_S^-+Eu_{Ca}^{(3+)}. \quad (3)$$

The resulting configuration can be represented as configuration (2), in which the Eu^{2+} and Eu^{3+} have changed places. Obviously, the Eu^{3+} ion here occupies an unfavorable position nearby the positively charged (relative to the lattice) center $[Cl_S^-]^{(+)}$ and, consequently, the given configuration has a large energy in comparison with configuration (2). If all the complexes (1) transition to this configuration, then precisely 50% of the Eu ions will wind up in the 2+ charge state. On the other hand, at high temperatures 47% of the europium is found in the form Eu^{2+} in configurations (1) and (2). Thus, it follows from our model that if almost all of the Eu is found in complexes (3) after illumination of the sample at low temperatures, then after thermal luminescence at high temperatures the amount of Eu^{2+} will hardly change. Such an effect was indeed observed in Ref. 2.

Free transition of an electron back from an Eu^{2+} ion to a $V_{Ca}^{(-)}$ vacancy does not take place, apparently because of the existence of a significant energy barrier (associated also with deformation of the complex). The process of release of the stored energy should start, in our opinion, with the capture of an electron by an Eu^{3+} ion:

$$Eu_{Ca}^{2+}-V_{Ca}^{(-)}-Cl_S^-+Eu_{Ca}^{(3+)}+e^- \rightarrow Eu_{Ca}^{2+}-V_{Ca}^{(-)}-Cl_S^-+Eu_{Ca}^{(2+)}. \quad (4)$$

In CaS:Eu,Cl the electrons located in traps at low temperatures begin to escape from them at $T=100-150$ K. Falling into europium complexes, they cause the number of Eu^{2+} ions to increase according to Eq. (4), which leads to the growth of the magnetic susceptibility and EPR line intensity observed in Ref. 2. As the temperature is raised further to 200–210 K, holes are liberated from the traps, thereby causing the complexes to luminesce in accordance with the equation

$$Eu_{Ca}^{2+}-V_{Ca}^{(-)}-Cl_S^-+Eu_{Ca}^{(2+)}+h^+ \rightarrow Eu_{Ca}^{3+}-V_{Ca}^{(-)}-Cl_S^-+Eu_{Ca}^{(2+)}+h\nu. \quad (5)$$

This luminescence takes place on the Eu^{2+} ion neighboring a Cl_S^- substituted site, to which the energy should be resonantly transferred from the Eu^{3+} ion, which is found in an excited state after capture of a hole. The possibility of resonant energy transfer from Eu^{3+} to Eu^{2+} also has experimental confirmation.[9] Finally, the complexes transition to configuration (2) and the number of Eu^{2+} ions and the associated magnetic susceptibility and EPR signal intensity decrease, in complete agreement with the experimental results of Ref. 2.

4. MAIN RESULTS AND CONCLUSIONS

We have carried out combined measurements of the magnetic susceptibility and EPR spectra of the small-crystal luminophors CaS:Eu,Cl. We found that:

Eu enters into the CaS matrix primarily in the trivalent state Eu^{3+}, and the fraction of ions in this state is 70–80% in the chlorine-free samples and increases with growth of the total europium concentration. Chlorine doping causes the Eu^{2+} fraction to increase to 50%.

The sample with low Eu concentration ($x=0.005$ at. %) compared with the residual Cl concentration possesses an anomalous temperature dependence of the susceptibility $\chi(T)$, which passes through a maximum at 5 K. Weak anomalies (which grow with chlorine doping) are also observed in other samples.

We have advanced the hypothesis that the anomalous temperature dependence of the susceptibility $\chi(T)$ is associated with recharging of Eu^{2+} to Eu^{3+}. We have proposed a structure of the complexes into which the europium ions enter: $Eu_{Ca}^{3+}-V_{Ca}^{(2-)}-Cl_S^-Eu_{Ca}^{3+}$ (1) and $Eu_{Ca}^{3+}-V_{Ca}^{(-)}-Cl_S^-Eu_{Ca}^{2+}$ (2), where complex (1) possesses zero magnetic moment and does not contribute to the magnetic susceptibility or the EPR signal.

The anomalous dependence of $\chi(T)$ is explained by recharging of europium in these complexes:

$$Eu_{Ca}^{3+} - V_{Ca}^{(2-)} - Cl_S^- - Eu_{Ca}^{(3+)} \rightarrow Eu_{Ca}^{3+} - V_{Ca}^{(-)} - Cl_S^- - Eu_{Ca}^{(2+)}.$$

On the basis of the proposed model, by fitting to the experimental data we have estimated the energy difference of complexes (1) and (2) to be $\Delta E \approx 13$ K.

We have shown that the proposed model of the complexes enables one to explain the mechanism of storage and release (luminescence) of light energy:

$$Eu_{Ca}^{3+} - V_{Ca}^{(2-)} - Cl_S^- - Eu_{Ca}^{(3+)} + h\nu \rightarrow Eu_{Ca}^{2+} - V_{Ca}^{(-)} - Cl_S^- - Eu_{Ca}^{(3+)},$$

$$Eu_{Ca}^{2+} - V_{Ca}^{(-)} - Cl_S^- - Eu_{Ca}^{(3+)} + e^- \rightarrow Eu_{Ca}^{2+} - V_{Ca}^{(-)} - Cl_S^- - Eu_{Ca}^{(2+)},$$

$$Eu_{Ca}^{2+} - V_{Ca}^{(-)} - Cl_S^- - Eu_{Ca}^{(2+)} + h^+ \rightarrow Eu_{Ca}^{3+} - V_{Ca}^{(-)} - Cl_S^- - Eu_{Ca}^{(2+)} + h\nu.$$

Additional studies are necessary to clarify the extent to which the given mechanism controls thermal luminescence for samples with different Eu and Cl content.

[1] R. S. Title, Phys. Rev. Lett. **3**, 273 (1959).
[2] M.-L. Yu. Allsadu, M. I. Danilkin, V. D. Kuznetsov *et al.*, Vestnik MGU, Ser. Fiz. **44**, 97 (1989).
[3] V. D. Kuznetsov, Prib. Tekh. Eksp. **No. 4**, 196 (1985).
[4] M. I. Danilkin, M. Must, E. U. Pedak *et al.*, Radiat. Meas. **24**, 351 (1985).
[5] É. L. Nagaev, *Physics of Magnetic Semiconductors* [in Russian], Nauka, Moscow (1979).
[6] A. A. Samokhvalov, N. V. Vol'kenshteĭn, T. D. Zotov *et al.*, Zh. Éksp. Teor. Fiz. **54**, 1341 (1968) [Sov. Phys. JETP **27**, 718 (1968)].
[7] V. T. Kalinnikov and Yu. V. Rakitin, *Introduction to Magnetochemistry* [in Russian], Nauka, Moscow (1980).
[8] T. M. Pham, N. Ruelle, E. Tronc *et al.*, Jpn. J. Appl. Phys. **33**, 1876 (1994).
[9] Sh. Fukumoto, Y. Hayashi, and S. Ibuki, Jpn. J. Appl. Phys., Part 2 **32**, L791 (1993).

Translated by Paul F. Schippnick

Nonmonotonic behavior of the superconducting transition temperature in bimetallic ferromagnet–superconductor structures

Yu. N. Proshin and M. G. Khusainov*)

Kazan State University, 420008 Kazan, Russia
(Submitted 7 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 1708–1730 (May 1998)

For layered ferromagnet/superconductor (F/S) structures we develop a theory of the proximity effect. In contrast to previous approaches, this theory allows for a finite transmission coefficient of the interface between the two metals and competition between the diffusion and spin-wave types of quasiparticle motion in the ferromagnet's strong exchange field. The superconductivity in F/S systems proves to be a superposition of BCS pairing with a constant-sign pair amplitude in the S-layers and Larkin–Ovchinnikov–Fulde–Ferrell (LOFF) pairing with an oscillating wave function in the F-layers. We show that the oscillatory behavior of the superconducting transition temperature T_c is due to oscillations of the Cooper pair flux from the S-layer to the F-layer, which are the result of oscillations of the discontinuity (jump) of the pair amplitude at the F/S boundary as the thickness d_f of the F-layer increases. The presence of nonmagnetic impurities leads to heavy damping of the oscillations of the LOFF pair amplitude and rapid deterioration of the coherent coupling of the boundaries of the F-layer in which the T_s vs. d_f dependence reaches a plateau as d_f increases. In F/S superlattices, in contrast to F/S double-layer junctions, there are two forms of the LOFF state, the 0-phase and the π-phase, which differ in their symmetry with respect to the center of the F layer. This gives rise to additional oscillations in the T_c (d_f) dependence due to the $0-\pi$ transitions. As the most vivid manifestation of LOFF states in F/S-systems, we predict the existence of recurrent and periodically recurrent superconductivities. We give a qualitative explanation of the different behavior of the superconducting transition temperature observed by different groups of experimenters dealing with the same ferromagnet–superconductor structures.
 [S1063-7761(98)01105-6]

1. INTRODUCTION

The antagonism that exists between superconductivity and ferromagnetism leads to a number of new nontrivial phenomena related to their coexistence and mutual adjustment in F/S junctions and superlattices formed by alternating ferromagnetic (F) and superconducting (S) layers.[1–14] For instance, the emergence of superconducting phases with incommensurate magnetic ordering in ferromagnetic insulator/superconductor structures[2,3] explains the presence of a nonuniform internal field that causes the BCS peak in the quasiparticle density states in EuO/Al (Ref. 4) and EuS/Al (Ref. 5) junctions to split, and the anomalously weak suppression of superconductivity in EuO/V multilayers.[6] New critical behavior of interphase boundaries and the spin-wave spectrum near an unusual Lifshitz point are expected in such systems.[3]

Bimetallic F/S structures also exhibit some unusual physical properties. The results of an analysis of recent experiments[7–11] with ferromagnet/superconductor multilayers suggest that the behavior of the dependence of the superconducting transition temperature T_c as a function of the thickness d_f of the ferromagnetic layers may be very different for the same F/S structures. In particular, while in some experiments involving Fe/V (Ref. 7) and Gd/Nb (Ref. 8) systems instead of the initial rapid drop in T_c with increasing d_f T_c reaches a plateau; in other experiments with the same systems (Ref. 9 and Refs. 10 and 11, respectively), the temperature T_c oscillates before reaching the plateau. The theoretical interpretation of oscillations in $T_c(d_f)$ in Ref. 12 and 13 reduces to periodic ''switching'' of the superconductor type from the 0-phase to the π-phase, with the sign of the order parameter Δ reversed as an F-layer is crossed. However, we will see that both theories (Refs. 12 and 13) are valid only when the transparence of the F/S boundary is high and are restricted by the extremely ''dirty'' limit of the ferromagnetic metal, with $2I\tau_f \ll 1$, where I is the exchange field and τ_f^{-1} is the collision rate of electrons and nonmagnetic impurities. Hence the limitations of Refs. 12 and 13 make it impossible to describe two different types of behavior of $T_c(d_f)$ in a unified manner. More than that, recently it has been found that oscillations of the function $T_c(d_f)$ also occur in the Fe/Nb/Fe triple-layer structure,[14] where π-phase superconductivity is impossible in principle. All this requires stating the problem of the nature of the nonmonotonic T_c (d_f) dependence in F/S systems anew and building a theory that gives a meaningful explanation of the existing experimental data.

Historically, the effect of quantum oscillations of T_c as a function of the thickness of the nonsuperconducting layer of a semiconductor or semimetal first arose in the experiments

of Golyanov and Mikheeva.[15] Kagan and Dubovskiĭ[16] showed this effect to be related to oscillations of the electron density of states in the superconducting metal, due to size quantization of the transverse motion of the current carriers in the nonmetallic film. However, when the normal metal is a ferromagnetic metal with a strong exchange field, there is another mechanism for the oscillations of $T_c(d_f)$, which we believe to be more suitable for an F/S structure and which is examined below.

Indeed, in systems such as Fe/V and Gd/Nd the ferromagnetic layers are strong magnets. The depairing action of the exchange polarization I is so strong that the presence of one atomic plane of magnetically ordered iron is enough to destroy the superconductivity of one hundred atomic planes of vanadium.[9] Hence an important condition for the preservation of superconductivity in such structures is moderately high transparency of the F/S boundaries (in addition to the requirement that the thickness d_s of the S-layers be large, or $d_s \gg d_f$). Probably, this condition corresponds to the experimental situation (see the discussion in Secs. 4 and 5). However, as noted by Radović *et al.*,[12] at present the microscopic boundary conditions for the wave function of Cooper pairs for arbitrary transparency of the F/S boundary are unknown. At the same time, the theory of the proximity effect in a normal metal/superconductor (N/S) system[17,18] suggests that the Cooper pair flux from the S-layer to the N-layeris proportional to the product of the transmission coefficient by the discontinuity (jump) of the pair amplitude at the S/N boundary. We can expect that boundary conditions of this type are valid for F/S systems, too.

Due to the proximity effect, pair correlations (to the extent to which the S/F boundary is transparent) are induced in F-layers by S-layers, but the large exchange splitting of the Fermi surface ($2I \gg \pi T_c$) changes the pairing conditions in the ferromagnet. In F-layers, pairs form quasiparticles from constant-energy states $\mathbf{p},\uparrow$ and $-\mathbf{p}+\mathbf{k},\downarrow$ whose momenta differ in absolute value, where $k \simeq 2I/v_f$ is the coherent momentum of the pair and v_f is the Fermi velocity. This is known as the Larkin–Ovchinnikov–Fulde–Ferrell (LOFF) state,[19,20] which is characterized by a pair amplitude oscillating with a period $a_f \simeq k^{-1}$. Scattering by nonmagnetic impurities, which has no effect on BCS pairing with a zero total momentum in S-layers, hinders the formation of a LOFF phase in F-layers.[21,22] Hence pair amplitude oscillations in relatively pure ferromagnetic layers with $2I\tau_f > 1$ decay over distances of order the mean free path $l_f = v_f\tau_f$ ($l_f > a_f$) from the S/F boundary. At the same time, it is known that the wave function of Cooper pairs in S-layers has a constant sign but an arbitrary phase. Thus, we should expect that superconductivity in F/S systems is a combination of BCS pairing in S-layers and LOFF pairing[19,20] in F-layers.

The essence of the proposed mechanism of oscillations of $T_c(d_f)$ can easily be understood if we turn to the example of an F/S double-layer junction. The condition that there be no LOFF pair flux across the external ferromagnet–vacuum boundary leads to oscillations of the discontinuity of the pair amplitude at the F/S boundary as the thickness of the F-layer increases. Each time a node of the LOFF pair amplitude finds itself at the F/S boundary, the discontinuity and the related Cooper pair flux from the S-layer to the F-layer are at their maximum values. Since on penetration of the F-layer the Cooper pairs are immediately destroyed by the strong exchange field, these values of the thickness of the F-layer correspond to minima in the T_c (d_f) dependence or even to disappearance of superconductivity. But when a crest of the LOFF pair amplitude finds itself at the F/S boundary, the Cooper pair flux across the S/F boundary is at its minimum. Such thicknesses of the F-layercorrespond to maxima in the T_c (d_f) dependence. This mechanism explains not only the oscillations of the superconducting transition temperature in F/S multilayers[10,11,14] but also new effects, such as a periodically recurrent superconductivity. Thus, the oscillations of the discontinuity of the pair amplitude lead to a distinctive periodic self-locking of the F/S boundary and hence to periodic compensation of the paramagnetic effect of the exchange field for the S-layers.

As the concentration of the nonmagnetic impurities in the F-layers rises, the spin-wave nature of quasiparticle motion, inherent in a pure ferromagnet, is replaced, for $2I\tau_f < 1$ ($l_f < a_f$), by the diffusion type.[23–25] In the process, the pair momentum k becomes a poor quantum number, and the heavily damped oscillations of the LOFF-state pair amplitude cease to ensure coherent coupling between the boundaries of a F-layer. In this case the T_c (d_f) dependence in F/S systems may become monotonic, a fact observed in the experiments of Koorevaar *et al.*[7] and Strunk *et al.*[8]

In this paper we develop a theory of the proximity effect for layered F/S systems. In contrast to previous theories,[12,13] this theory allows the F/S boundary to have a finite transmission coefficient and permits competition between the diffusion and spin-wave types of quasiparticle motion in the ferromagnet layers. Section 2 is devoted to a novel microscopic derivation of the Usadel equations and the corresponding boundary conditions, which relate the flux of the Usadel function to the discontinuity of the function at the F/S boundary. The solution of the resulting boundary-value problem for the proximity effect in the F/S junction, given in Sec. 3, makes it possible to establish the dependence of the superconducting transition temperature T_c on the parameter $2I\tau_f$, on the transmission coefficient σ_s of the F/S boundary, and on the thicknesses of the ferromagnetic (d_f) and superconducting (d_s) layers. In Sec. 4 we study the competition between the 0- and π-phase types of superconductivity in F/S superlattices. Finally, Sec. 5 is devoted to a discussion of the results.

2. THE PROXIMITY EFFECT IN AN INHOMOGENEOUS *F/S* SYSTEM: GENERAL FORMALISM

Near a second-order phase transition point, the superconducting transition (or critical) temperature T_c of an inhomogeneous superconductor can be found by solving the Gor'kov integral equation for the order parameter $\Delta(\mathbf{r})$ (Ref. 26):

$$\Delta(\mathbf{r}) = V(\mathbf{r}) T \sum_{\omega}{}' \int H(\mathbf{r},\mathbf{r}',\omega)\Delta(\mathbf{r}')d^3r', \qquad (1)$$

where $V(\mathbf{r})$ is the electron–electron interaction potential at point specified by the radius vector $\mathbf{r}$, the prime on the summation sign indicates cutoff at the Debye frequency ω_D, $\omega=\pi T(2n+1)$ is the Fermi frequency, with $n=0,\pm1,\pm2,\ldots$, and T is the temperature. The kernel $H(\mathbf{r},\mathbf{r}',\omega)$ of the integral equation (1) is given by the following expression:

$$H(\mathbf{r},\mathbf{r}',\omega)=\frac{1}{2}\,\mathrm{Tr}\langle\hat{g}^{-1}\hat{G}(\mathbf{r},\mathbf{r}',\omega)\hat{g}\hat{G}(\mathbf{r},\mathbf{r}',-\omega)\rangle, \qquad (2)$$

where $\hat{G}(\mathbf{r},\mathbf{r}',\omega)$ is the matrix Green's function of electrons in the normal phase, $\hat{g}=i\sigma_y$, with σ_y the Pauli matrix, and the angle brackets stand for averaging over all the impurity configurations.

Let us consider a planar junction of a ferromagnetic metal (F) occupying the region $-\infty<z<0$ and a superconductor (S) occupying the region $0<z<\infty$. Due to the translation invariance of the F/S junction in the xy plane, the order parameter Δ depends only on z and Eq. (1) becomes

$$\Delta(z)=\frac{1}{2}V(z)T\sum_{\omega}{}'\sum_{\alpha\neq\beta}\int_{-\infty}^{\infty}H_{\alpha\beta}(z,z',\omega)\Delta(z')dz', \qquad (3)$$

where $V(z>0)=V_s$, $V(z<0)=V_f$, α and β are spin indices ($\alpha\neq\beta$), and

$$H_{\alpha\beta}(z,z',\omega)=\int d^2\rho H_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega),$$

where $\rho=(\mathbf{r}-\mathbf{r}')_\perp$ is the two-dimensional radius vector in the junction plane. Diagrammatic methods[27] reveal that in the presence of an exchange field and nonmagnetic scattering by impurities, the kernel $H_{\alpha\beta}(z,z',\omega)$ of Eq. (3) is the solution of another integral equation,

$$H_{\alpha\beta}(z,z',\omega)=K_{\alpha\beta}(z,z',\omega)+\int_{-\infty}^{\infty}\frac{K_{\alpha\beta}(z,z_1,\omega)H_{\alpha\beta}(z_1,z',\omega)}{2\pi N(z_1)\tau(z_1)}dz_1, \qquad (4)$$

where we have introduced the following notation:

$$K_{\alpha\beta}(z,z',\omega)=\int\frac{d^2p}{(2\pi)^2}\tilde{K}_{\alpha\beta}(\mathbf{p},z,z',\omega),$$

$$\tilde{K}_{\alpha\beta}(\mathbf{p},z,z',\omega)=G_{\alpha\alpha}(\mathbf{p},z,z',\omega)G_{\beta\beta}(\mathbf{p},z,z',-\omega), \qquad (5)$$

$G_{\alpha\alpha}(\mathbf{p},z,z',\omega)$ is the Green's function of the conduction electron in the normal phase averaged over the impurity configurations, $\mathbf{p}$ is the two-dimensional momentum in the junction plane, and $N(z)$ and $\tau^{-1}(z)$ are, respectively, the density of states on the Fermi level and the rate of scattering by impurities (both change their values suddenly when the boundary $z=0$ is crossed). Solving the problem with a potential barrier at the boundary between two semi-infinite metals in the spirit of Ref. 28, we can show that $\tilde{K}_{\alpha\beta}(\mathbf{p},z,z',\omega)$ obeys the differential equation

$$\left[2\omega_{\alpha\beta}(z)-v_z(z)l_{\alpha\beta}(z)\frac{\partial}{\partial z^2}\right]v_z(z)\tilde{K}_{\alpha\beta}(z,z')=2\delta(z-z') \qquad (6)$$

with the boundary conditions

$$v_{sz}l_{s\alpha\beta}\frac{\partial\tilde{K}_{\alpha\beta}(z,z')}{\partial z}\bigg|_{z=+0}=v_{fz}l_{f\alpha\beta}\frac{\partial\tilde{K}_{\alpha\beta}(z,z')}{\partial z}\bigg|_{z=-0}=\frac{\sigma}{2(1-\sigma)}[v_{sz}\tilde{K}_{\alpha\beta}(+0,\,z')-v_{fz}\tilde{K}_{\alpha\beta}(-0,\,z')], \qquad (7)$$

where for brevity some arguments have been dropped, v_z is the z-component of the electron velocity, σ is the barrier transmission coefficient, and

$$\omega_{\alpha\beta}(z)=|\omega|+\frac{1}{2\tau(z)}+iI(z)g_{\alpha\beta}\mathrm{sgn}\;\omega,$$
$$l_{\alpha\beta}(z)=\frac{v_z(z)}{2\omega_{\alpha\beta}(z)}. \qquad (8)$$

Here the dependence of all quantities on z is steplike, i.e., $I(z>0)=0$ and $I(z<0)=I$, $v_z(z>0)=v_{sz}$ and $v_z(z<0)=v_{fz}$, and $\tau(z>0)=\tau_s$ and $\tau(z<0)=\tau_f$.

By using Eq. (6) with the boundary conditions (7) we can reduce the problem of solving the integral equation (4) for an inhomogeneous F/S-system to an integro-differential boundary-value problem for the quantity $\tilde{H}_{\alpha\beta}(\mathbf{p},z,z',\omega)$, which is related to the kernel $H_{\alpha\beta}(z,z',\omega)$ by the formula

$$H_{\alpha\beta}(z,z',\omega)=\int\frac{d^2p}{(2\pi)^2}\tilde{H}_{\alpha\beta}(\mathbf{p},z,z',\omega). \qquad (9)$$

Note that due to the proximity of the characteristic momenta to the Fermi momenta, integration with respect to $\mathbf{p}$ in (5) and (9) reduces to integration with respect to the angle ϑ between the electron velocity vector $\mathbf{v}$ and a unit vector $\mathbf{n}$ normal to the boundary:

$$\int\frac{d^2p}{(2\pi)^2}\to\pi N(z)\int_0^1 v_z(z)dx=\pi N(z)v(z)\int_0^1 x\;dx, \qquad (10)$$
$$x=\cos\;\vartheta.$$

Plugging (4) and (5) into (6) and (7), we arrive at the following integro-differential equation:

$$\left[2\omega_{\alpha\beta}(z)-v_z(z)l_{\alpha\beta}(z)\frac{\partial^2}{\partial z^2}\right]\pi N(z)v_z(z)\tilde{H}_{\alpha\beta}(\mathbf{p},z,z',\omega)=\frac{1}{\tau(z)}H_{\alpha\beta}(z,z',\omega)+2\pi N(z)\delta(z-z') \qquad (11)$$

with boundary conditions for $\tilde{H}_{\alpha\beta}(\mathbf{p},z,z',\omega)$ of the form

$$v_{sz}l_{s\alpha\beta}\frac{\partial\tilde{H}_{\alpha\beta}(z,z')}{\partial z}\bigg|_{z=+0}=v_{fz}l_{f\alpha\beta}\frac{\partial\tilde{H}_{\alpha\beta}(z,z')}{\partial z}\bigg|_{z=-0}$$
$$=\frac{\sigma}{2(1-\sigma)}[v_{sz}\tilde{H}_{\alpha\beta}(+0,\ z')$$
$$-v_{fz}\tilde{H}_{\alpha\beta}(-0,\ z')]. \quad (12)$$

If we solve the boundary-value problem represented by the system of equations (9)–(12), we can find the kernel $H_{\alpha\beta}(z,z',\omega)$. Then, substituting the kernel into he integral equation, we can find (in principle) the superconducting transition temperature T_c and the spatial dispersion of the order parameter $\Delta(z)$ of the inhomogeneous F/S system without imposing restrictions on the value of the barrier transmission coefficient σ and the purity of the ferromagnetic and superconducting materials.

The above integro-differential boundary-value problem (9)–(12) for the proximity effect in an F/S junction simplifies considerably in the dirty limit, where the electron mean free path $l_j=v_j\tau_j$ $(j=f,s)$ becomes much smaller than

$$\xi_{j\omega}=\mathrm{Re}\sqrt{\frac{D_j}{2(|\omega|+iI_j)}},$$

the range of the kernel $H_{\alpha\beta}(z,z',\omega)$ (here $D_j=v_jl_j/3$ is the diffusion coefficient). We also assume that l_f is smaller than the spin rigidity length $a_f=v_f/2I$ of the ferromagnet, with $2I\tau_f<1$. An analysis of Eqs. (9)–(11) suggests that the anisotropic correlator $v_z(z)\tilde{H}_{\alpha\beta}(\mathbf{p},z,z',\omega)$ can be written as a sum of two terms,

$$\pi N(z)v_z(z)\tilde{H}_{\alpha\beta}(\mathbf{p},z,z',\omega)=H_{\alpha\beta}(z,z',\omega)$$
$$+h_{\alpha\beta}(\mathbf{p},z,z',\omega), \quad (13)$$

an isotropic term $H_{\alpha\beta}(z,z',\omega)$ with a characteristic scale $\xi_{j\omega}$, and an anisotropic term $h_{\alpha\beta}(\mathbf{p},z,z',\omega)$ with a range of order l_j $(\ll\xi_{j\omega})$ (the anisotropic term vanished after angle-averaging). The representation (13) reflects the physical fact that the two-particle Green's function $v_z(z)\tilde{H}_{\alpha\beta}(\mathbf{p},z,z',\omega)$ becomes more and more isotropic as the impurity concentration rises (see, e.g., Ref. 29).

Since the spatial scales of interest to us are large compared to the mean free paths l_j, we can ignore $h_{\alpha\beta}$ in (13) everywhere except the region near the F/S boundary. This corresponds to what is known as the diffusion approximation.[30] Inserting (13) in (11), averaging the result over the angle between the velocity vector and the z axis, and taking into account what was said earlier, we arrive at the following differential equation of diffusion form for the kernel $H_{a\beta}(z,z',\omega)$:

$$\left[|\omega|+iI(z)g_{\alpha\beta}\ \mathrm{sgn}\ \omega-\frac{1}{2}D_{\alpha\beta}(z)\frac{\partial^2}{\partial z^2}\right]H_{\alpha\beta}(z,z',\omega)$$
$$=\pi N(z)\delta(z-z'). \quad (14)$$

Note that this equation is related to Eq. (11) in the same way as the Usadel equations are related to the Eilenberger equations (see, e.g., Ref. 29). Since Eq. (14) does not describe the real behavior of $H_{\alpha\beta}(z,z',\omega)$ near the boundary between the metals, it must be augmented by appropriate boundary condition, which asymptotically smooth out at distances greater than l_j. The required expressions that link the flux of the kernel $H_{\alpha\beta}(z,z',\omega)$ across the junction to the value of the kernel at the F- and S-sides of the junction can be derived directly from the exact boundary conditions (12). In the case of a flat F/S boundary, which conserves the component of momentum parallel to the surface of the junction, these boundary conditions are

$$D_{s\alpha\beta}\frac{\partial H_{\alpha\beta}(z,z')}{\partial z}\bigg|_{z=+0}=D_{f\alpha\beta}\frac{\partial H_{\alpha\beta}(z,z')}{\partial z}\bigg|_{z=-0}$$
$$=\frac{1}{4}[\sigma_sv_sH_{\alpha\beta}(+0,z')$$
$$-\sigma_fv_fH_{\alpha\beta}(-0,z')], \quad (15)$$

where σ_s and σ_f are the transmission coefficients at the junction from the S- and F-sides, respectively. The two transmission coefficients are related by the detailed balance condition $\sigma_sv_sN_s=\sigma_fv_fN_f$ (see Ref. 17), which states that the numbers of transitions from the S-layer to the F-layer and back are equal. In (14) and (15) we have introduced the following notation:

$$D_{\alpha\beta}(z)=\frac{D(z)}{1+2i\tau(z)I(z)g_{\alpha\beta}\ \mathrm{sgn}\ \omega},$$
$$D(z)=D_j, \quad \sigma_j=\left\langle\frac{\sigma v_{jz}}{(1-\sigma)v_j}\right\rangle. \quad (16)$$

The complex-valued denominator in $D_{\alpha\beta}(z)$ describes the suppression of the diffusive motion of the conduction electrons in the ferromagnet by the strong exchange field of the localized spins.[23] Because of this the motion of the spin carriers in the ferromagnet acquires a mixed diffusion–spin-wave nature. As the parameter $2I\tau_f$ increases, the spin-wave nature becomes dominant. The angle brackets $\langle\cdots\rangle$ denote averaging over the angle between the direction of electron velocity and a normal to the junction surface. Equations (3), (14), and (15) generalize the problem of the proximity effect for an arbitrary transmission coefficient of the N/S boundary[17,18] to the case where the normal metal (N) is a ferromagnet with an exchange splitting $2I$ that is considerably smaller than the Fermi energy E_f. In this case the differences in the densities of states and the transmission coefficients of the F/S boundary for electrons with oppositely directed spins can be ignored.

If we define an anomalous Usadel function $F_{\alpha\beta}(z,\omega)$ (see Refs. 12,13, and 29) as

$$F_{\alpha\beta}(z,\omega)=\frac{1}{\pi N(z)}\int_{-\infty}^{\infty}H_{\alpha\beta}(z,z',\omega)\Delta(z')dz', \quad (17)$$

Eqs. (14) with the boundary conditions (15) can easily be transformed into the corresponding boundary-value problem for the Usadel functions. In terms of the functions $F_{\alpha\beta}(z,\omega)$, the problem of the proximity effect in an F/S junction is described by an equation for the order parameter obtained from (3),

$$\Delta(z)=\frac{1}{2}\lambda(z)\pi T\sum_{\alpha\neq\beta}\sum_{\omega}{}' F_{\alpha\beta}(z,\omega) \tag{18}$$

($\lambda(z)=N(z)V(z)$ is the dimensionless electron–electron interaction parameter), and Usadel equations that are linearized for $T\to T_c$,

$$\left[|\omega|+iI(z)g_{\alpha\beta}\ \mathrm{sgn}\ \omega-\frac{1}{2}D_{\alpha\beta}(z)\frac{\partial^2}{\partial z^2}\right]F_{\alpha\beta}(z,\omega)=\Delta(z), \tag{19}$$

with boundary conditions that relate the flux of the pair amplitude $F_{\alpha\beta}(z,\omega)$ to the discontinuity of the amplitude at the boundary $z=0$:

$$D_{s\alpha\beta}\frac{\partial F_{\alpha\beta}(z,\omega)}{\partial z}\bigg|_{z=+0}=\frac{\sigma_s v_s}{4}[F_{\alpha\beta}(+0,\omega)-F_{\alpha\beta}(-0,\omega)],$$
$$D_{f\alpha\beta}\frac{\partial F_{\alpha\beta}(z,\omega)}{\partial z}\bigg|_{z=-0}=\frac{\sigma_f v_f}{4}[F_{\alpha\beta}(+0,\omega)-F_{\alpha\beta}(-0,\omega)]. \tag{20}$$

The resulting boundary-value problem (18)–(20) for the proximity effect in the F/S junction differs from the previous boundary-value problems (see Refs. 12 and 13) in two ways. First, the boundary conditions used in Refs. 12 and 13 and amounting to the continuity of the Usadel function in the passage through the plane $z=0$ constitute a particular case of Eqs. (20) and correspond to the limit of a large transmission coefficient, $\sigma_j\gg l_j/\xi_{j\omega}$, i.e., the flux of $F_{\alpha\beta}(z,\omega)$ across the F/S boundary can be ignored. But σ_j, which determines the amplitude of the Cooper pair flux from the S-layer to the F-layer,[17] strongly depends on the conditions and method of formation of the F/S boundary. Hence either it can serve as am adjustable parameter or must be measured experimentally. Second, suppression of the diffusive motion of the spin carriers by the ferromagnet′s exchange field results in the emergence of spin rigidity,[23–25] i.e., the appearance of an imaginary part in the effective diffusion coefficient (see Eq. (16) and the beginning of Sec. 3):

$$D_f(I)=\frac{D_f}{1+2iI\tau_f}. \tag{21}$$

Because of this, the solutions of the Usadel equation (19) in a sufficiently pure F-layer with a strong exchange field oscillate with relatively low damping, which suggest that there is a LOFF state.[19,20] Oscillations of the discontinuity in the pair amplitude at the F/S boundary lead, in accordance with the boundary conditions (20), to periodic variation in the Cooper pair flux from the S-layer to the F-layer, thus forcing oscillations to appear in the $T_c(d_f)$ dependence. These oscillations, whose observability strongly depends on σ_j, prove to be possible not only in F/S multilayers but also in F/S double layers.

3. SUPERCONDUCTING TRANSITION TEMPERATURE OF AN *FIS* DOUBLE-LAYER JUNCTION

As an example of how the obtained Usadel equations (19) and the boundary conditions (20) can be used, we calculate the superconducting transition temperature of a planar F/S junction in which the ferromagnetic metal occupies the region $-d_f<z<0$ and the superconductor, the region $0<z<d_s$. Allowing for the symmetry properties of the Usadel function, $F_{\alpha\beta}(z,\omega)=F^*_{\beta\alpha}(z,\omega)=F_{\beta\alpha}(z,-\omega)$ and going over to positive frequencies ω, we write Eqs. (18) and (19) separately for the S- and F-layers, discarding the spin indices for convenience. For the F-layer we obtain

$$\Delta_s(z)=2\lambda_s\pi T\ \mathrm{Re}\sum_{\omega>0}{}' F_s(z,\omega), \tag{22}$$

$$\left[\omega-\frac{D_s}{2}\frac{\partial^2}{\partial z^2}\right]F_s(z,\omega)=\Delta_s(z),\quad 0<z<d_s. \tag{23}$$

Similarly, for the F-layer we have

$$\Delta_f(z)=2\lambda_f\pi T\ \mathrm{Re}\sum_{\omega>0}{}' F_f(z,\omega), \tag{24}$$

$$\left[\omega+iI-\frac{D_f(I)}{2}\frac{\partial^2}{\partial z^2}\right]F_f(z,\omega)=\Delta_f(z),\quad -d_f<z<0, \tag{25}$$

with $D_f(I)$ specified in (21). The boundary conditions at the plane $z=0$ corresponding to Eqs. (23) and (25) become

$$D_s\frac{\partial F_s(z,\omega)}{\partial z}\bigg|_{z=+0}=\frac{\sigma_s v_s}{4}[F_s(+0,\omega)-F_f(-0,\omega)],$$
$$D_f(I)\frac{\partial F_f(z,\omega)}{\partial z}\bigg|_{z=-0}=\frac{\sigma_f v_f}{4}[F_s(+0,\omega)-F_f(-0,\omega)]. \tag{26}$$

We seek the solutions to Eqs. (23) and (25) together with boundary conditions (26) and the conditions

$$\frac{\partial F_{f,s}}{\partial z}(z=-d_f,d_s)=0, \tag{27}$$

corresponding to a zero electron flux across the external boundaries of the junction in the form ($\omega>0$)

$$F_s(z,\omega)=\frac{\Delta_{s0}}{\omega}+A\frac{\cosh[q_s(z-d_s)]}{\cosh\, q_s d_s},\quad 0<z<d_s,$$
$$F_f(z,\omega)=\frac{\Delta_{f0}}{\omega+iI}+B\frac{\cosh[q_f(z+d_f)]}{\cosh\, q_f d_f},\quad -d_f<z<0, \tag{28}$$

where Δ_{s0} and Δ_{f0} are the initial values of the order parameters for the thin S- and F-layers, A and B are parameters determined from the conditions (26) and (27), and the wave numbers q_s and q_f are given by the following expressions:

$$q_s^2=\frac{2\omega}{D_s},\quad q_f^2=\frac{2(\omega+iI)}{D_f(I)}. \tag{29}$$

The mutual effect of the metals F and S across the boundary is especially significant in the Cooper limit, where the thicknesses d_j of the layers are small compared to the coherence lengths

$$\xi_j=\mathrm{Re}\sqrt{\frac{D_j(I_j)}{2(\pi T+iI_j)}},\quad j=f,s;\quad I_f=I,\quad I_s=0.$$

In this case the order parameter and the Usadel function are essentially constant inside the F- and S-layers, since $|q_j d_j| \ll 1$ in (28). Then the condition that Eqs. (22) and (24) are simultaneously nontrivial yields, after we sum over the frequencies ω, the following expression for the reduced superconducting transition temperature $t = T_c/T_{cs}$ of the F/S junction (T_{cs} is the transition temperature of the isolated S-layer):

$$[\ln t - \mathrm{Re}(c_s'\chi_- + c_f'\chi_+)][\ln t + \lambda_f^{-1} - \lambda_s^{-1} - \mathrm{Re}(c_f'\chi_- + c_s'\chi_+)] - [\mathrm{Re}\, c_s''(\chi_- - \chi_+)][\mathrm{Re}\, c_f''(\chi_- - \chi_+)] = 0. \tag{30}$$

Here we have introduced the following notation:

$$\chi_\pm = \Psi\left(\frac{1}{2}\right) - \Psi\left(\frac{1}{2} + \frac{\Gamma_\pm}{2\pi T_{cs} t}\right) + \ln\left(1 + \frac{\Gamma_\pm}{\omega_D}\right),$$

$$\Gamma_\pm = \frac{\Gamma + iI \pm \sqrt{(\Gamma + iI)^2 - 4iI\Gamma_s}}{2},$$

$$\Gamma = \Gamma_s + \Gamma_f, \quad \Gamma_j = \frac{\sigma_j v_j}{8 d_j}, \tag{31}$$

$$c_f' = \frac{\Gamma_+ - \Gamma_f - iI}{\Gamma_+ - \Gamma_-}, \quad c_s' = \frac{\Gamma_+ - \Gamma_s}{\Gamma_+ - \Gamma_-},$$

$$c_f'' = \frac{\Gamma_s}{\Gamma_+ - \Gamma_-}, \quad c_s'' = \frac{\Gamma_f}{\Gamma_+ - \Gamma_-},$$

where $\Psi(x)$ is the digamma function, and Γ_f and Γ_s are the electron transition rates from the F-layer and S-layer and back, respectively. At $I = 0$ Eq. (30) becomes the equation for the superconducting transition temperature of the N/S junction at arbitrary values of the transmission coefficient of the interface of the metals, which was derived earlier in Ref. 17 by one of the present authors. Here we have $\Gamma_- = 0$ and $\Gamma_+ = 0$, and the coefficients $c_j' = c_j''$ are equal to c_j, where

$$c_j = \frac{N_j d_j}{N_s d_s + N_f d_f}, \quad c_f \Gamma_f = c_s \Gamma_s, \tag{32}$$

i.e., c_f and c_s are relative bulk densities of electron states in the metals F and S, respectively.

Here we have examined only the corollaries of Eq. (30) that reflect the experimental facts, according to which $I \gg \pi T_{cs}$ and $d_s \gg d_f$. In particular, the suppression of T_c for the smallest thicknesses d_f of the ferromagnetic layer is described by the formula

$$\ln t \simeq -\frac{c_f(\lambda_s - \lambda_f)}{\lambda_s(c_s\lambda_s + c_f\lambda_f)} + \Psi\left(\frac{1}{2}\right) - \mathrm{Re}\, \Psi\left(\frac{1}{2} + \frac{i c_f I}{2\pi T_{cs} t}\right), \quad \Gamma_f > I. \tag{33}$$

This formula predicts strong initial lowering of T_c with increasing d_f because of the averaging of the electron–electron interactions λ_f and λ_f and also because of spin polarization of I over the entire sample due to the rapid electron exchange between the F- and S-layers. The exchange rate is characterized by the parameter $\Gamma = \Gamma_f + \Gamma_s$ (see Ref. 17), where in this case $\Gamma_f > I > \omega_D \gg \Gamma_s, \pi T_{cs}$. Here, initially for $c_f I \ll \pi T_{cs}$, the suppression of T_c by the first term in (33) dominates, since the initial contribution of the exchange field is quadratic in the parameter c_f and $|\lambda_f| \ll \lambda_s$ probably holds due to the electron repulsion (supplementing attraction) via spin waves in the ferromagnet.[31] Then, as d_f increases, a more rapid drop in T_c ensures the depairing action of the average exchange polarization $c_f I$. For large thicknesses of the F-layer, when Γ_f becomes smaller than I, the F/S boundary becomes effectively self-locked and the F and S subsystems are weakly coupled, so that instead of Eq. (33) for T_c, Eq. (30) yields

$$\ln t \simeq \Psi\left(\frac{1}{2}\right) - \Psi\left(\frac{1}{2} + \frac{\Gamma_s}{2\pi T_{cs} t}\right), \quad \Gamma_f < I. \tag{34}$$

In this case the rate Γ_s at which pairs leave the superconductor for the ferromagnet acts as the depairing parameter, and for $\Gamma_s < \pi T_{cs}/2\gamma$ ($\gamma = 1.781$ is Euler's constant) the function $T_c(d_f)$ becomes a finite constant. Hence, for a given thickness of the F-layer, superconductivity in the F/S junction arises only if d_s is larger than the critical thickness $d_{sc} = \sigma_s v_s \gamma/4\pi T_{cs} \simeq \sigma_s \xi_{s0}$, where ξ_{s0} is the BCS coherence length. The relation $\Gamma_f \simeq I$ allows for a minimum (for small transmission coefficients) estimate of the critical thickness of the F-layer, $d_{fc} \simeq \sigma_f v_f/8I = \sigma_f a_f$, above which T_c ceases to depend on d_f. The existence of the critical thicknesses d_{sc} and d_{fc} has been noted in all experiments with F/S multilayers.[7–11,14]

The possibility of the Usadel function in the F-layer behaving in an oscillatory manner and the experimental conditions require considering arbitrary thicknesses d_f and d_s. As demonstrated earlier, the difference in the electron–electron interaction parameters ($\lambda_f < \lambda_s$) causes the superconducting transition temperature of the F/S junction to decrease only up to thicknesses d_f of order the interatomic distance. The strong depairing effect of the exchange field I ($I \gg \pi T_{cs}$) is the main mechanism for the destruction of superconductivity in F/S systems. Hence, ignoring (to make matters simpler) the order parameter induced in the F-layer, $\Delta_f(z) \simeq 0$ ($\lambda_f \simeq 0$), we seek the solutions of Eqs. (22)–(27) in the form

$$F_s(z,\omega) = \frac{\Delta_s(z)}{\omega + D_s k_s^2/2} = \frac{C}{\omega + D_s k_s^2/2} \times \frac{\cos[k_s(z - d_s)]}{\cos k_s d_s}, \quad 0 < z < d_s,$$

$$F_f(z,\omega) = E\frac{\cos[k_f(z + d_f)]}{\cos k_f d_f}, \quad -d_f < z < 0, \tag{35}$$

where C and E are constants, and k_s and k_f are wave numbers independent of the frequency ω ($\omega \ll I$). Then for the superconducting transition temperature of the F/S junction we obtain an ordinary equation of the Abrikosov–Gor′kov type,

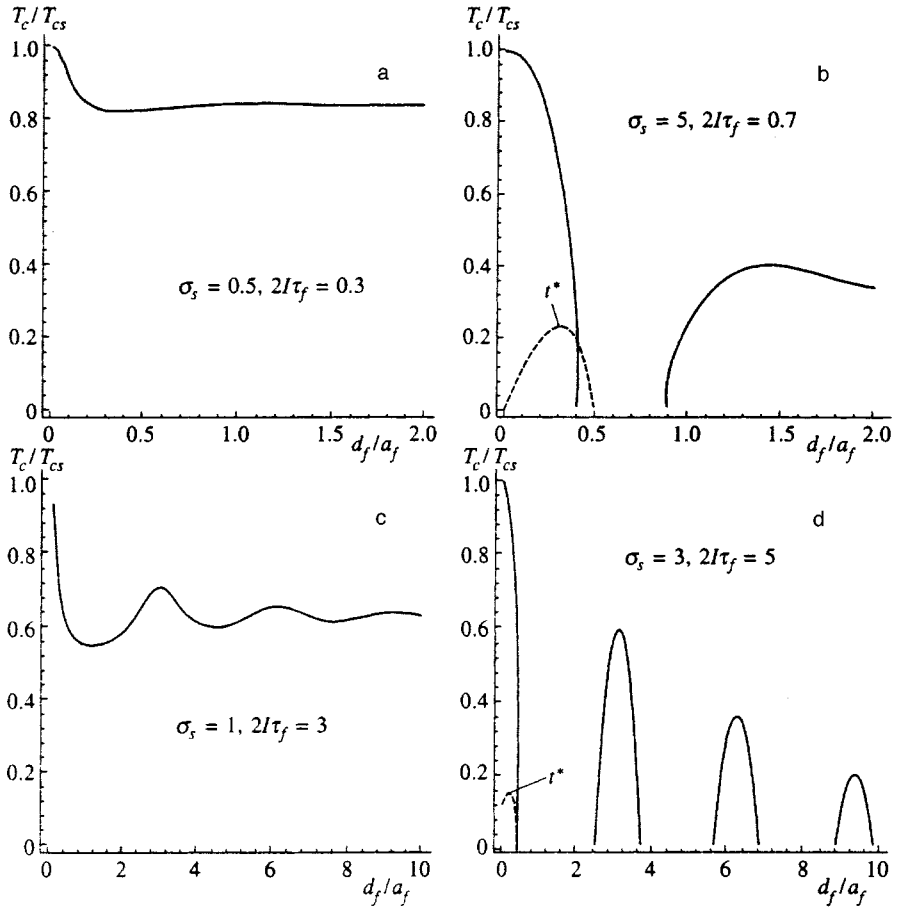

FIG. 1. Reduced superconducting transition temperature T_c/T_{cs} of the F/S junction vs. the reduced thickness d_f/a_f of the ferromagnetic layer, described by Eqs. (36), (37), and (41) at $N_s v_s = N_f v_f$, $d_s = 500$ Å, and $\xi_{s0} = 400$ Å, where ξ_{s0} is the BCS coherence length. The values of the parameters σ_s and $2I\tau_f$ are given in each diagram; (a) T_c reaches a plateau, (b) recurrent superconductivity, (c) oscillations of T_c, and (d) periodically recurrent superconductivity. The dashed curves labeled t^* in Figs. b and d are lines of tricritical points.

$$\ln t = \Psi\left(\frac{1}{2}\right) - \mathrm{Re}\ \Psi\left(\frac{1}{2} + \frac{D_s k_s^2}{4\pi T_{cs} t}\right), \tag{36}$$

where the depairing parameter $D_s k_s^2$ is the solution of another equation,

$$D_s k_s \tan k_s d_s = \frac{\sigma_s v_s}{4 - [\sigma_f v_f / D_f(I) k_f] \cot k_f d_f}, \tag{37}$$

and the wave number k_f for $\tau_f^{-1} > 2I \gg 2\pi T_{cs}$ is given by the following expression:

$$k_f^2 = \frac{-2iI}{D_f(I)} = \frac{-2iI(1+2iI\tau_f)}{D_f}. \tag{38}$$

Plugging the solutions (35) into the first boundary condition in (20), we can easily see that the left-hand side of Eq. (37), which defines the depairing parameter $D_s k_s^2$ in Eq. (36) for T_c, is proportional to the Cooper pair flux from the S-layer to the F-layer. Here the resonant denominator of the right-hand side of Eq. (37), being inversely proportional to the discontinuity in the pair amplitude at the F/S boundary, periodically changes the value of this flux as the thickness of the F-layer increases due to the function $\cot k_f d_f$.

In the Cooper limit $|k_j d_j| < 1$, the system (36)–(38) reproduces formulas similar to (33) for $\Gamma_f > I$ and (34) for $\Gamma_f < I$. Hence the single-mode approximation[32] used in deriving Eq. (36), which is an approximation in which multiple reflection and the passage of electron waves across the F/S boundary are ignored, is not too crude. Going beyond the Cooper limit, where the oscillatory nature of the pair amplitude in the F-layer cannot clearly manifest itself, requires a more detailed analysis of Eqs. (36)–(38).

Spatial variations of the Usadel function in the F-layer are characterized by the wave number $k_f = 1/\xi_f' - i/\xi_f''$, where ξ_f' is responsible for the oscillation period and ξ_f'', for the depth of penetration of the pair amplitude into the ferromagnet. In the dirty limit ($l_f \ll a_f \ll v_f/2\pi T$) these two lengths are approximately equal and, as Eq. (38) suggests, differ only in small corrections of order $I\tau_f$:

$$\xi_f' = \sqrt{D_f/I}(1 - I\tau_f), \quad \xi_f'' = \sqrt{D_f/I}(1 + I\tau_f), \quad 2I\tau_f \ll 1. \tag{39}$$

In this case, due to the heavy damping of the function $F_f(z,\omega)$, as d_f increases, only one node of this function, which leads to a minimum in the T_c (d_f) dependence, can reach the F/S boundary. As d_f increases further, the coherent coupling of the two boundaries of the F-layer deteriorates and the Cooper pair flux from the S-layer to the F-layer becomes constant, with the T_c (d_f) dependences reaching a plateau. For small values of the parameter σ_s and $2I\tau_f$, the minimum of T_c almost merges with the plateau, as Fig. 1a clearly shows. Such behavior, with the function $T_c(d_f)$ first rapidly decreasing and then reaching a plateau, was observed in bimetallic Gd/Nb junctions[9] and is also known from early experiments with F/S double layers.[33] However, when the transmission coefficient σ_s of the F/S boundary is large, the development of this minimum in T_c with increasing $2I\tau_f$

may lead to recurrent superconductivity, depicted in Fig. 1b. Clearly, there are two critical thicknesses here: d_{fc1}, at which superconductivity in the F/S junction disappears, and d_{fc2}, at which superconductivity is restored in the system. Hence Eq. (37) can be interpreted as a relation from which we can find all the crystal thicknesses mentioned earlier: d_{fc} (at larger thicknesses T_c is independent of d_f), d_{sc} (at smaller thicknesses the superconductivity in the F/S junction is suppressed), and d_{fc1} and d_{fc2} as functions of the exchange field I and the transmission coefficient σ_s of the FS boundary.

When the transmission coefficient σ_s of the FS boundary is large and the exchange field is strong (as in the case depicted in Fig. 1b), we must allow for a possible change of the phase transition type. The tricritical point t_c^* below which the transition from a superconducting state to a normal state is second-order, can be found by simultaneously solving Eq. (36) and the equation for the line of tricritical points $t^*=T^*/T_{cs}$:

$$\mathrm{Re}\ \Psi^{(2)}\left(\frac{1}{2}+\frac{D_s k_s^2}{4\pi T_{cs} t^*}\right)=0, \tag{40}$$

where $\Psi^{(2)}(x)$ is the tetragamma function. The region where the $T_c(d_f)$ curve intersects the line of tricritical points t^* is characterized by the fact that the solution of Eq. (36) for T_c in this region is two-valued, so that new states of the LOFF type[19,20] can occur in this region. Probably, a new LOFF state appears via additional sinusoidal modulation of the pair amplitude in the xy plane of the junction with a characteristic wave vector $\mathbf{q}_\perp=\mathbf{i}q_x+\mathbf{j}q_y$. Here, due to an increase in the coherence length $\xi_s=\sqrt{D_s/2\pi T}$, at low temperatures modulation extends not only to the F-layer but also to the S-layer. Hence the neighborhood of the critical point t_c^* in the T_c (d_f) dependence requires a special investigation, which lies outside the scope of the present paper.

If $2I\tau_f>1$ holds (this is possible, since $I\sim 10^3$K for Gd and Fe), the diffusion approximation used in our description of conduction electron motion in the ferromagnetic is inapplicable,[23] although the ordinary condition for the dirty limit, $\pi T\tau_f\ll 1$, may be met. The important thing is that the mean free path l_f becomes longer than the spin rigidity length a_f, which is now the smallest characteristic scale in the problem. In this case, quasiparticle motion in the ferromagnet is predominantly spin-wave, and scattering by impurities leads to slowly decaying spin waves.

Note that our approach to the proximity effect in F/S systems, developed in Sec. 2, is generally free (in contrast to the previous theories[12,13]) from restrictions on the strength of impurity scattering in the F- and S-layers. In particular, for $\pi T\tau_i\ll 1$ but arbitrary $2I\tau_f$, there is no way in which Eq. (25) for $F_f(z,\omega)$ can be used, although Eq. (23) for $F_s(z,\omega)$ in this case remains valid. In this case, by solving Eqs. (9), (11), and (17) simultaneously, we can show that the characteristic equation for k_f is more complicated than Eq. (38):

$$\frac{k_f l_f}{\arctan[k_f l_f/(1+2iI\tau_f)]}=1. \tag{41}$$

An approximate solution of this equation for $2I\tau_f<1$ corresponds to the adopted diffusion approximation and yields formula (38) for k_f.

If $2I\tau_f>1$ holds, the asymptotic behavior of $F_f(z,\omega)$ is described by a wave number k_f that differs somewhat from (38):

$$k_f^2\approx\frac{-2iI(1+2iI\tau_f)}{v_f l_f}, \tag{42}$$

where $\mathrm{Re}\ k_f$ is a quantity of order the coherent pair momentum k in a LOFF state,[19,20] and $\mathrm{Im}k_f\sim l_f^{-1}$ determines the spread in the values of k, i.e., the decay of the LOFF phase. Here the oscillation period of the pair amplitude, ξ_f', becomes much smaller than the depth of penetration of pairs into the ferromagnet, ξ_f'':

$$\xi_f'\approx a_f=\frac{v_f}{2I},\quad \xi_f''\simeq 2l_f,\quad 2I\tau_f\gg 1. \tag{43}$$

To within the $D_f\to 3D_f$ substitution, the system of equations (36)–(38) proves applicable in a qualitative description of the $T_c(d_f)$ dependence in F/S junctions with relatively pure ferromagnetic layers, too. As d_s increases, the Usadel function, the depairing factor $D_s k_s^2$, and the $T_c(d_f)$ dependence oscillates with a period determined by the spin rigidity length a_f. These oscillations decay for $d_f>2l_f$ $(\gg a_f)$, causing $F_f(z,\omega)$ to vanish as we move away from the F/S boundary and making T_c constant, as shown in Fig. 1c. Such behavior of $T_c(d_f)$ was observed in the Fe/Nb/Fe triple-layer junction.[14] To generalize our results to $F/S/F$ triple layers, we need only replace d_s with $d_s/2$ in the above formulas. Interestingly, for sufficiently large values of the parameters σ_s and $2I\tau_f$, the superconductivity of the F/S junction at low temperature is of a periodically recurrent nature, with sections that are superconductors within the F-layer alternating with normal sections, as shown in Fig. 1d.

Clearly, the maxima and minima in the $T_c(d_f)$ dependence in Figs. 1c and 1d appear when, respectively, crests and nodes of the pair amplitude $F_f(z)\propto\cos[k_f(z+d_f)]$ cross the F/S boundary $(z=0)$. The Cooper pair flux from the S-layer to the F-layer, given by Eq. (37), decreases $(\sin k_f d_f\to 0)$ or increases $(\cos k_f d_f\to 0)$ in step with the discontinuity of the pair amplitude at the F/S boundary. Formally, this looks like periodic modulation of the transmission coefficient of the F/S boundary or like periodic compensation of the paramagnetic effect of the exchange field of the F-layer. As the impurity concentration in the F-layer increases to $\tau_f^{-1}>2I$, the motion of paired particles becomes diffusive and, according to (39), $\mathrm{Re}\ k_f\simeq\mathrm{Im}\ k_f\sim\sqrt{l_f a_f}$. Then the oscillations of the pair amplitude in the LOFF phase become heavily damped, and their effect on the behavior of $T_c(d_f)$ is much weaker than when $2I\tau_f>1$ holds (Figs. 1a and 1b).

4. COMPETITION OF 0-PHASE AND π-PHASE STATES IN F/S MULTILAYERS

To calculate the superconducting transition temperature of an F/S superlattice formed by alternating ferromagnetic layers of thickness d_f and superconducting layers of thick-

ness d_s, we study the spatial variations of the Usadel function in a unit cell consisting of an F-layer ($-d_f<z<0$) and an S-layer ($0<z<d_s$). To this end we augment the system of equations (22)–(26) by periodicity conditions for the Usadel function and the order parameter:

$$F_{s,f}(z+L,\omega)=e^{i\varphi}F_{s,f}(z,\omega), \quad \Delta(z+L)=e^{i\varphi}\Delta(z), \tag{44}$$

where $L=d_f+d_s$ is the superlattice period, and φ is an arbitrary phase. In addition, we require that boundary conditions of type (26) are met at the outer boundaries $z=-d_f,d_s$ of the unit cell of the superlattice. In what follows we assume that the interaction of the magnetization of the neighboring F-layers through the superconducting S-layers can be ignored. This assumption is justified for F/S multilayers in which the thickness d_s of the S-layers is larger than ξ_s, since in this case the indirect RKKY exchange between neighboring F-layers is exponentially small (see, e.g., Refs. 2 and 3). This corresponds to the conditions of the experiments discussed in Refs. 7–11, and 14, where it was noted that the critical thicknesses d_{sc} of the superconducting layers below which there can be no superconductivity in the F/S system are larger than ξ_s. In these cases the mutual orientation of the magnetizations of the neighboring F-layers in metallic F/S superlattices is unimportant.

Clearly, all the physically distinct values of the phase φ lie within the segment $0\leqslant\varphi\leqslant\pi$. However, the general solution of the boundary-value problem (22)–(26), (44) implies that the only the solutions with $\varphi=0$ and $\varphi=\pi$ are the most stable in a broad range of thicknesses d_f and d_s. Estimates have shown[12,13] that the ranges where the intermediate values of the phase $0<\varphi<\pi$ are realized are extremely narrow and correspond to transitions from one stable state to another. Hence here we give only the solutions for the 0-phase ($\varphi=0$) and π-phase($\varphi=\pi$) states of the F/S superlattice.

In the 0-phase case, the characteristic solutions of our boundary-value problem for the Usadel function are those with crests at the centers of the S- and F-layers:

$$F_s(z,\omega)\propto\cos[k_s(z-d_s/2)], \quad 0<z<d_s, \tag{45}$$

$$F_f(z,\omega)\propto\cos[k_f(z+d_f/2)], \quad -d_f<z<0. \tag{46}$$

These solutions lead to the already familiar equation (36) for the superconducting transition temperature T_c, where the depairing factor $D_sk_s^2$ is now given by the equation

$$D_sk_s\tan\frac{k_sd_s}{2}=\frac{\sigma_sv_s}{4-[\sigma_fv_f/D_f(I)k_f]\cot(k_fd_f/2)}, \tag{47}$$

which differs from the similar equation (37) for the F/S junction in that d_j is replaced by $d_j/2$. This is a natural consequence of the symmetry of the superlattice if one ignores the interaction between neighboring F-layers through the superconducting layers.

In the π-phase case, the corresponding solution for the Usadel function has a node at the center of the F-layer (in contrast to (46)), i.e.,

$$F_f(z,\omega)\propto\sin[k_f(z+d_f/2)], \quad -d_f<z<0. \tag{48}$$

Hence, the depairing parameter $D_sk_s^2$ in Eq. (36) for T_c is described by a different transcendental equation:

$$D_sk_s\tan\frac{k_sd_s}{2}=\frac{\sigma_sv_s}{4+[\sigma_fv_f/D_f(I)k_f]\tan(k_fd_f/2)}. \tag{49}$$

Note that the only way in which the right-hand side of Eq. (47) differs from the right-hand side of Eq. (49) is that the second contains $-\tan x$ where the first contains $\cot x$. The wave number k_f for both equations, Eq. (47) (the 0-phase) and Eq. (49) (the π-phase), is determined by the expression (38) for $2I\tau_f<1$ and (42) for $2I\tau_f>1$. We also note that the 0-phase (46) and π-phase (48) solutions, differing in the parity of the pair amplitude with respect to the center of the F-layer, are two variants of the quasi-one-dimensional LOFF state[19,20] in F/S superlattices. Here the pair amplitude is doubly periodic: ''microscopically'' inside each F-layer and ''macroscopically'' in the superlattice as a whole. The latter period in the 0-phase case is $L=d_f+d_s$, while in the 0-phase case it is $2L$. A comparison of (47) and (49) shows that now, as d_f increases, the Cooper pair flux may also oscillate due to transitions between 0- and π-phase states.

The results of a numerical analysis of the behavior of $T_c(d_f)$ for the 0- and π-phase states of an F/S superlattice for different values of the parameters are depicted in Fig. 2. Figure 2a shows the $T_c(d_f)$ dependence, from which it follows that for $\sigma_s\simeq2I\tau_f\simeq0.5$ the transition from the 0-phase branch to the π-phase branch leads to a single burst in the superconducting transition temperature, which then reaches a plateau. This behavior of $T_c(d_f)$ resembles the results of experiments[10,11] with Gd/Nb multilayers. For somewhat smaller values of these quantities ($\sigma_s\sim2I\tau_f\sim0.3$) the 0- and π-phase solutions in the ''switching'' region come closer, and instead of the burst in T_c in the same Gd/Nb multilayers[8] and in Fe/V superlattices,[7] the superconducting transition temperature reaches a plateau. Our analysis shows that the differences in the nature of the $T_c(d_f)$ dependence reported by different groups of researchers for the same F/S systems can probably be explained by the different transmission coefficients of the F/S boundaries and the different purity of the ferromagnetic layers. This in turn may be related to differences in the experimental conditions and the methods of preparation of F/S structures used by the different groups. We also note that when the transmission coefficients of the boundaries are large in F/S superlattices with $2I\tau_f<1$, recurrent superconductivity is possible (see Fig. 2b). In contrast to F/S junctions (Fig. 1b), recurrent superconductivity in F/S superlattices is characterized by the proximity, in the $T_c(d_f)$ phase diagram, of the 0- and π-phase regions limited by the thickness of the F-layer. The first time at which T_c vanishes is when 0-phase nodes approach F/S boundaries. The finite value of T_c is restored when the nodes of the π-phase pair amplitude approaches the boundaries. The second time T_c vanishes is when π-phase nodes approach F/S boundaries. A further increase in d_f leads to a situation in which the heavy damping of both 0- and π-phase solutions for $2I\tau_f<1$ destroys the coherent coupling between the neighboring F/S boundaries in the superlattice and the Cooper pair flux becomes constant but sufficiently strong (in this case) to suppress superconductivity.

Figure 2 shows that for relatively pure F-layers with a

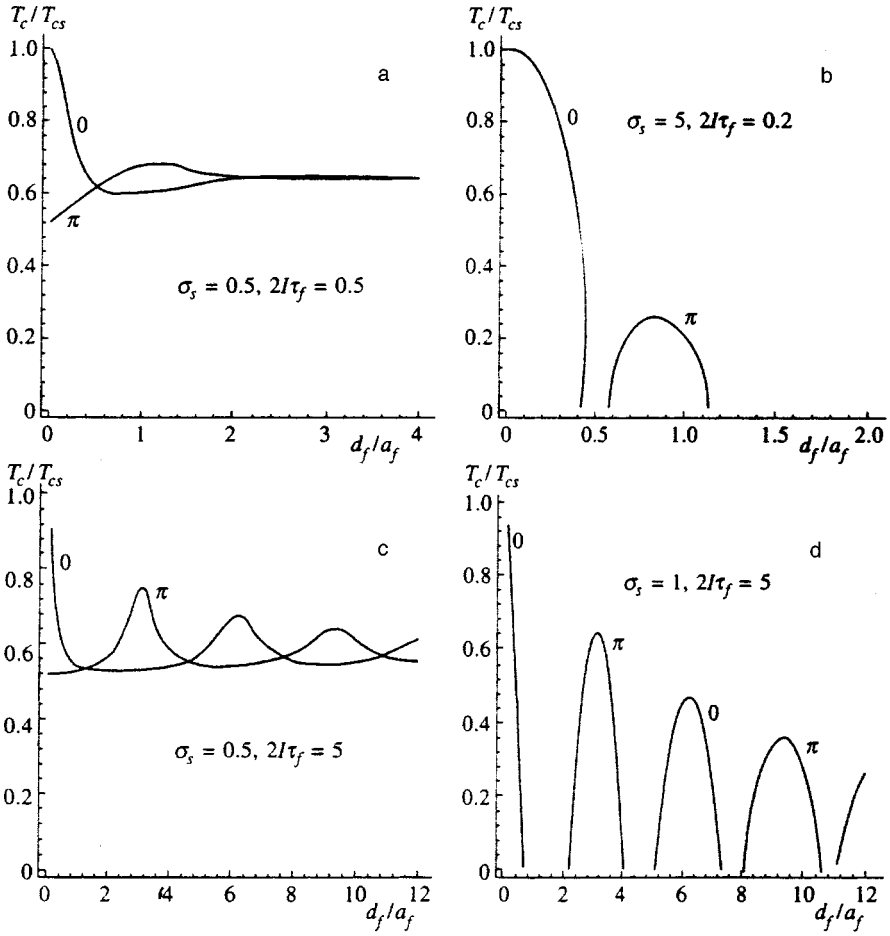

FIG. 2. The $(T_c/T_{cs}, d_f/a_f)$ phase diagrams of the F/S superlattice for $N_s v_s = N_f v_f$, $d_s = 500$ Å, and $\xi_{s0} = 400$ Å, where ξ_{s0} is the BCS coherence length. The values of the parameters σ_s and $2I\tau_f$ are given in each diagram; (a) single burst in T_c, (b) recurrent $0-\pi$ superconductivity, (c) oscillations of T_c with competition of 0- and π-phase states, and (d) periodically recurrent superconductivity with alternation of 0- and π-phase states. The 0 and π labeling the curves indicate that the curves belong to 0- and π-phase states, respectively.

strong exchange field ($2I\tau_f > 1$) and moderate transmission coefficients of the F/S boundaries, the competition of 0- and π-phase states causes them to alternate in the $T_c(d_f)$ phase diagram. Here the oscillation periods for the function $T_c(d_f)$ for 0- and π-phase solutions are twice as large, separately, as the periods in the corresponding F/S junction. The competition of the same states leads to a situation in which the oscillation periods of $T_c(d_f)$ in junctions and in superlattices coincide. Periodically recurrent superconductivity, which is possible when both σ_s and $2I\tau_f$ are large, is depicted in Fig. 2d. In F/S superlattices it emerges as an alternation of 0- and π-phase peaks separated by nonsuperconducting sections in the phase diagram. The points at which the 0- and π-phase branches of the $T_s(d_f)$ dependence intersect the lines of tricritical points t^* (to simplify matters we do not show then here) correspond, as in the case of F/S junctions (Fig. 1d), to the points at which the second-order phase transition becomes first-order. Near these points one should expect the emergence of new states of the LOFF type,[19,20] which are probably realized (just as they are in F/S junctions) through additional sinusoidal modulation of the pair amplitude in the xy plane. However, the phases of these two-dimensionally modulated structures in the neighboring S-layers coincide in the 0-phase case, but probably differ by π in the π-phase case.

5. CONCLUSIONS AND DISCUSSION

Here is a short list summarizing the results of our studies.

1. We have proposed a model of the superconducting state in F/S systems that allows for a combination of BCS pairing with zero total momentum in S-layers and LOFF pairing[19,20] with a coherent momentum of transverse motion, $k \simeq 2I/v_f$, in F-layers. The pair amplitude in S-layers has a constant sign, while in F-layers it oscillates with a period $a_f = k^{-1}$. Scattering by nonmagnetic impurities for $2I\tau_f < 1$ gives rise to heavy damping as the distance from the F/S boundary grows and to loss of coherent coupling between the two boundaries of an F-layer.

2. The oscillatory dependence of the superconducting transition temperature on the thickness of the F-layers is related to the oscillations of the discontinuity in the pair amplitude at the F/S amplitude, which lead to oscillations in the Cooper pair flux from the S-layer to the F-layer. This nonmonotonic dependence is inherent not only in multilayers but also in F/S double layers, which makes it possible (in contrast to the earlier theories of Radović *et al.*[12] and Buzdin *et al.*[13]) to explain the presence of oscillations in the $T_c(d_f)$ dependence in the Fe/Nb/Fe triple-layer structure.[14]

3. As the most striking consequence of the quasi-one-dimensional realization of LOFF states, we have predicted

the phenomena of recurrent and periodically recurrent superconductivities in F/S junctions and superlattices.

4. We have established that, depending on the value of the transmission coefficient of the F/S boundary and the purity of the ferromagnetic layers in F/S multilayers of the same composition, the $T_s(d_f)$ dependence may reach a plateau in a monotonic manner or in an oscillatory manner. This provides an explanation for the contradiction between the results of different groups of experimenters (see Ref. 8 and Refs. 10 and 11 for Gd/Nb multilayers and Ref. 7 and Ref. 9 for Fe/V superlattices).

The difference of the Usadel equation (25) stated in the present paper from a similar equation obtained in earlier theories[12,13] refers only to the ferromagnetic metal and arises because it was essentially postulated rather than derived. The approach, discussed in Sec. 2, to deriving Usadel equations was based on using the integral equation (4) for the correlator $H_{\alpha\beta}(z,z',\omega)$. This approach makes it possible to obtain not only Eq. (19) for the pair amplitude $F(z,\omega)$ but also the necessary boundary conditions (20). The reason is that the integral equation (4) contains the full information about the parameters of the electronic structure and the transport characteristics of the metals comprising the junction as well as the information about the abrupt variation of the respective quantities where the well-defined F/S boundary is crossed. The boundary conditions (20) relating the flux of the Usadel function to the discontinuity at the F/S boundary generalize the conditions obtained earlier in Ref. 34 for a junction comprised of dirty normal and superconducting metals to the case of an F/S junction. Strictly speaking, of the boundary conditions used in Refs. 12 and 13 only the first is correct, i.e., the one that states that the fluxes from F to S and back are proportional. The second condition concerning the continuity of the pair amplitude, $F(+0,\omega)=F(-0,\omega)$, introduced by de Gennes,[35] is valid only for $\sigma_j \to \infty$ $(\sigma=1)$. For a finite value of the transmission coefficient of the F/S boundary, the function $F(z,\omega)$ has a discontinuity proportional to the Cooper pair flux across the boundary, as follows from (20) and as noted in Refs. 32 and 34. This fact together with the oscillatory behavior of the pair amplitude in F-layers causes the superconducting transition temperature to oscillate in F/S systems. Clearly, the transmission coefficient σ_j, which controls the amplitude of the Cooper pair flux from the S-layer to the F-layer, is largely determined by the technology of preparing F/S structures. Hence it must either be found from independent experiments (say, from electrical conductivity measurements) or be a adjustable parameter in the theory. Indeed, our results, reflected in Eqs. (37), (47), and (49) and in Figs. 1 and 2, suggest that together with $2I\tau_f$ the parameter σ_j plays an extremely important role in describing the various types of $T_c(d_f)$ dependence in layered F/S systems. The above formulas also suggest that decreasing the transparency of the F/S boundary (say, by moderate oxidation of the boundary in the process of its formation) makes it possible to preserve the superconducting properties of the system even when the S-layers are very thin $(d_s<\xi_s)$.

The difference in the nature of the mutual adjustment of superconductivity and ferromagnetism in superlattices and in F/S junctions is due primarily to the nonequivalence of the boundaries of the ferromagnetic layer in the latter. Indeed, the condition that there is no flux of LOFF pairs across the ferromagnet–vacuum boundary pins a crest of the function $F_f(z,\omega)$ at this boundary, and an increase in the thickness d_f of the F-layer leads to the above-mentioned oscillations in the discontinuity of the pair amplitude at the F/S boundary and in the Cooper pair flux across the boundary, and to a nonmonotonic $T_c(d_f)$ dependence. Because of their periodicity along the z axis, F/S superlattices with a period $L=d_f+d_s$ have equivalent boundaries within one unit cell. This means that they allow not only for cosine (0-phase) solutions but also for sine (π-phase solutions, which pin a crest and node of the pair amplitude at the center of an F-layer. In accordance with the theory of second-order phase transitions, within which we operate, in the given conditions a state with a higher superconducting transition temperature (a lower free energy) is realized. Hence the phase diagrams represented in Fig. 2 describe different cases of competition of 0- and π-phase states due to simultaneous node–crest transitions at neighboring F/S boundaries of a superlattice with increasing d_f.

From the formal viewpoint, it can be expected that any mechanism leading to oscillations in the discontinuity of the pair amplitude and the Cooper pair flux across the boundary separating a superconductor S and a normal material also leads to oscillations of T_c as the thickness of the normal material increases. In this sense, the mechanism proposed by Kagan and Dubovskiĭ[16] that explains the oscillations in T_c by the existence of quantum coupling between the two boundaries of a semimetal or a semiconductor due to oscillations of the electron wave function, is universal. However, if the normal part of the junction is a metal, the reflection of the electron waves with a wavelength on the order of the interatomic distance from inevitable roughnesses of the surface of the same scale complete averages these oscillations. In the presence of impurities, this averaging is due to the scattering of electrons in the bulk of the film if the film is thicker than the mean free path. Hence, in the case where a ferromagnetic metal with a strong exchange field is in contact with a superconductor, we believe that the oscillatory nature of the pair amplitude in the F-layer is related to a LOFF state,[19,20] and the paramagnetic effect is, naturally, the mechanism by which Cooper pairs arriving from the S-layer are destroyed. Here one must bear in mind that the absence of electron–electron attraction in the F-layer ($\lambda_f=0$ in our model) does not stop quasiparticles from forming pairs. In view of the proximity effect, pair correlations, whose source is the S-layer, are induced in the F-layer up to depths on the order of the coherence length ξ_f''. Indeed, as shown in Sec. 3, electron exchange, whose intensity strongly dependence on the transparency of the boundary, causes collectivization of electron–electron interactions in the junction layer with a thickness of order $\xi_s+\xi_f''$. We note in this connection that the true order parameter for our inhomogeneous F/S system is not $\Delta(z)=\lambda(z)F(z)/N(z)$ but the pair amplitude[35]

$$F(z)\propto \mathrm{Re}\sum_{\omega>0} F(z,\omega),$$

which, in contrast to Δ, does not vanish when we go over to a F-layer.

Since all the results obtained in this paper are interpreted in terms of realization of a LOFF state with an oscillating pair amplitude in F-layers, we will briefly discuss the nontrivial effect of impurity scattering on this phase. As is well known, a LOFF state is extremely sensitive the presence of nonmagnetic impurities.[21,22] The reason is that pair correlations in an F-layer are effective only for quasiparticles, whose constant-energy states $(\mathbf{p},\uparrow;-\mathbf{p}+\mathbf{k},\downarrow)$ are not mutually inverted in time, and the Anderson theorem is invalid for them. It is no surprise, then, that ξ_f'' acts as the coherence length in an F-layer; for $2I\tau_f>1$ this length is of order of the mean free path l_f. Scattering by nonmagnetic impurities leads to damping of the oscillations of the function $F_f(z,\omega)$ as we move deeper into the F-layer and reduces the amplitude of this function at the F/S boundary at $z=0$. This is why T_c at the maximum points ($d_f\propto\pi a_f$) is restored only partially and not to its nominal value T_{cs} (see Figs. 1c, 1d, 2c, and 2d). At the first maximum point the value of the depairing parameter $D_s k_s^2$ proves to be of order $d_f/d_s\tau_f$. As the impurity concentration rises to $2I\tau_f<1$, the coherence length and the oscillation period of $F_f(z,\omega)$ decrease to $\xi_f''\simeq\xi_f'\sim\sqrt{a_f l_f}$. In this case, in the T_c (d_f) dependence there can be only one minimum (node) in F/S junctions and two minima (0 and π) in superlattices (see Figs. 1a, 1b, 2a, and 2b). After this the connection between the boundaries of F-layers is lost and a constant Cooper pair flux from S-layers to F-layers sets in.

What is important is that the approach to the description of the proximity effect in F/S systems developed in Sec. 2 is free, in principle, from restrictions on the rate of impurity scattering, τ_j^{-1} $(j=f,s)$. However, the simple equations (19) for the pair amplitude can be obtained only in the dirty limit for $\pi T\tau_j\ll 1$ and $2I\tau_f\ll 1$. Thus, the dirty limit condition for a ferromagnetic metal with $I\gg\pi T$ proves to be much more stringent than for the superconducting part of the F/S sample, which limits the applicability of Eq. (25) for the function $F_f(z,\omega)$ considerably. For $\pi T\tau_f\ll 1$ but an arbitrary value of the parameter $2I\tau_f$, we are forced to seek a simultaneous solution of Eqs. (9), (11), and (17), which leads to the characteristic equation (47) for the complex-valued wave number k_f. This makes it possible to study the effect of competition of the diffusion $(2I\tau_f<1)$ and spin-wave $(2I\tau_f>1)$ types of quasiparticle motion on the spatial behavior of the pair amplitude in a ferromagnet. In such strong ferromagnets as Fe and Gd, the exchange splitting of the Fermi surface amounts to $2I\sim 10^3$K, so that the value of the parameter $2I\tau_f$ to a great extent depends on the purity of the metal. For instance, for Gd with $v_f=1.2\times 10^5\mathrm{ms}^{-1}$ and the mean free path $l_f\sim 20$ Å we have $2I\tau_f\sim 2$. The spin rigidity length a_f in this case is of order 10 Å, so that the oscillations of $T_c(d_f)$ should be appreciable. However, mutual diffusion of the metals and uncontrollable impurities introduced in the process of preparing the F/S structures may considerably reduce the value of $2I\tau_f$ (especially in thin F-layers) and make it impossible to detect the oscillations.

The presence of oscillations in the $T_c(d_f)$ dependence does not necessarily mean that the experimental phase diagram of superlattices contains a π-phase. The thing is that these oscillation occur in 0- and π-phase states, but their periods are twice as large as in the case of a F/S junction. The competition of these states related to the separation of the maxima in T_c by $\delta d_f=\pi a_f$ leads a situation in which the periods of oscillations in F/S junctions and superlattice formally become equal. Hence only simultaneous experiments with F/S double layers and multilayers can irrefutably prove the presence of a π-phase in the T_c (d_f) phase diagram. Note that, in contrast to the 0-phase state, the π-phase state is characterized not only by the presence of pair-amplitude oscillations inside the F-layer but also by the fact that the phase of this amplitude changes by π when any layer (F or S) is passed. In this sense the π-phase state having a one-dimensionally modulated and pair amplitude and an order parameter closely resembles the LOFF state[19,20] in ferromagnetic superconductors. As a LOFF state, the π-phase is a state with a spontaneous superconducting current. However, as Fulde and Ferrell showed in their pioneering paper,[20] in a LOFF phase this current is fully balanced by the current of unpaired quasiparticles. Probably, a similar situation emerges in π-phase F/S superlattices, which may serve as ideal model systems for studying states of the LOFF type.

All our results were obtained under the assumption that on the atomic scale the F/S boundary is well-defined and that ferromagnetic ordering exists even in the thinnest F-layers possible. Clearly, the situation is much more complicated for the real F/S systems studied in the experiments described in Refs. 7–11 and 14. Sometimes in view of the mutual solubility of the F and S metals and sometimes for purely technological reasons, the F/S boundary is found to be smeared. Hence thin layers of a ferromagnetic material may prove to be paramagnetic[8] or even nonmagnetic.[14] In principle, the adopted model of F/S systems can easily be broadened to incorporate the case of a thin nonmagnetic layer between the F- and S-layers or allow for a paramagnet–ferromagnet phase transition when the F-layer reaches a critical thickness. Physically it is clear, however, that if the depth of penetration ξ_f'' is much larger than the typical thickness of the nonferromagnetic layers (7–10 Å), oscillations in the Cooper pair flux through such a smeared S/F boundary will still remain. Probably, the effect of buffer layers with fairly thick F-layers can roughly be taken into account by a simple renormalization of the exchange field on the electron–electron interaction constant λ_s.

A critical comparison of the oscillatory T_c vs. d_f curves depicted in Figs. 1c, 2a, and 2c (for F/S junctions and superlattices) and the experimental results (Ref. 14 and Refs. 10 and 11, respectively) indicates two important discrepancies. First, only one maximum is detected in $T_c(d_f)$. Second, the experimental plateau for this function always lies below the first minimum. The first discrepancy can be satisfactorily explained by the contamination of the ferromagnetic layers by nonmagnetic impurities (so that $2I\tau_f\leqslant 1$ holds) and the slight or moderate transparency of the F/S boundary ($\sigma_s<1$). For superlattices this means that the curves represented in Fig. 2a must be replaced by the curves in Fig. 2a, while for F/S and $F/S/F$ functions this means that the second maximum in Fig. 1c is too low to be observed against the

plateau. The second discrepancy may be due primarily to the fact that the proposed theory does not allow for the effect of spin and spin–orbit scattering of electrons in the ferromagnetic layers. Incidentally, nonmagnetic impurities in the ferromagnetic act as spin scatterers.[26] Since spin and spin–orbit scattering mixes the quasiparticle spins $\uparrow$ and $\downarrow$, it will destroy LOFF pair correlations[21,22] and cause an additional decrease in T_c, this increasing the discontinuity in the pair amplitude at the F/S boundary. Here, in view of the proximity effect, spin scattering has a depairing effect also on the BCS condensate of the superconducting layer to depths of order ξ_s. Additional suppression of T_c by spin and spin–orbit scattering increases with the thickness of the F-layer up to values of order ξ_f''. Hence the positions of the minima and the point at which the $T_c(d_f)$ dependence reaches a plateau gradually become lower. It can also be expected that the spin and spin–orbit scattering will lead to additional damping of the pair amplitude in the F-layer, which may also have an effect on the observability of oscillations in the T_c (d_f) dependence. Further development of the proposed theory is advisable and will facilitate the construction of a model of F/S systems that correctly explains their nature.

We are grateful to I. A. Garifullin, Yu. A. Izyumov, M. Yu. Kagan, B. I. Kochelaev, G. B. Teĭtel'baum, and I. M. Suslov for the interest in the present work and for the useful discussions.

*)E-mail: mansur.khusainov@ksu.ru

[1] B. Y. Jin and J. B. Ketterson, Adv. Phys. **38**, 189 (1989).
[2] M. G. Khusainov, Zh. Éksp. Teor. Fiz **109**, 524 (1996) [JETP **82**, 278 (1996)].
[3] M. G. Khusainov, Zh. Éksp. Teor. Fiz **110**, 966 (1996) [JETP **83**, 533 (1996)].
[4] P. M. Tedrow, J. E. Tkaczyk, and A. Kumar, Phys. Rev. Lett. **56**, 1746 (1986).
[5] X. Hao, J. S. Moodera, and R. Meservey, Phys. Rev. Lett. **67**, 1342 (1991).
[6] G. M. Roesler, M. E. Filipkowski, P. R. Broussard *et al.*, in *Proc. SPIE Int'l. Soc. Opt. Eng. (USA)*, Vol. 2157 (1994), p. 285.
[7] P. Koorevaar, Y. Suzuki, R. Coehoorn, and J. Aarts, Phys. Rev. B **49**, 441 (1994).
[8] C. Strunk, C. Surgers, U. Paschen, and H. V. Lohneysen, Phys. Rev. B **49**, 4053 (1994).
[9] H. K. Wong, B. Y. Jin, H. Q. Yang, J. B. Ketterson, and J. E. Hilliard, J. Low Temp. Phys. **63**, 307 (1986).
[10] J. S. Jiang, D. Davidović, D. H. Reich, and C. L. Chien, Phys. Rev. Lett. **74**, 314 (1995).
[11] J. S. Jiang, D. Davidović, D. H. Reich, and C. L. Chien, Phys. Rev. B **54**, 6119 (1996).
[12] Z. Radović, M. Ledvij, L. Dobrosavljević-Gruvić, A. I. Buzdin, and J. R. Clem, Phys. Rev. B **44**, 759 (1991).
[13] A. I. Buzdin, B. Bujicic, and M. Yu. Kupriyanov, Zh. Éksp. Teor. Fiz. **101**, 231 (1992) [Sov. Phys. JETP **74**, 124 (1992)].
[14] Th. Mühge, N. N. Garif'yanov, Yu. V. Goryunov, G. G. Khaliullin, L. R. Tagirov, K. Westerholt, I. A. Garifullin, and H. Zabel, Phys. Rev. Lett. **77**, 857 (1996).
[15] V. M. Golyanov and M. N. Mikheeva, Zh. Éksp. Teor. Fiz. **70**, 2236 (1976) [Sov. Phys. JETP **43**, 1167 (1976)].
[16] Yu. Kagan and L. B. Dubovskiĭ, Zh. Éksp. Teor. Fiz. **72**, 646 (1977) [Sov. Phys. JETP **45**, 339 (1977)].
[17] M. G. Khusainov, JETP Lett. **53**, 579 (1991).
[18] A. A. Golubov, in *Proc. SPIE Int'l. Soc. Opt. Eng. (USA)*, Vol. 2157 (1994), p. 353.
[19] A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **47**, 1136 (1964) [Sov. Phys. JETP **20**, 762 (1965)].
[20] P. Fulde and R. A. Ferrell, Phys. Rev. **135**, A550 (1964).
[21] L. G. Aslamazov, Zh. Éksp. Teor. Fiz. **55**, 1477 (1968) [Sov. Phys. JETP **28**, 773 (1969)].
[22] S. Takada, Prog. Theor. Phys. **43**, 27 (1970).
[23] P. Fulde and A. Luther, Phys. Rev. **175**, 337 (1968).
[24] L. L. Hirst, Phys. Rev. **141**, 503 (1966).
[25] J. I. Kaplan, Phys. Rev. **143**, 351 (1966).
[26] Ø, H. Fischer, Helv. Phys. Acta **45**, 331 (1972).
[27] A. A. Abrikosov, L. P. Gor'kov, and I. Ye. Dzyaloshinski, *Quantum Field Theoretical Methods in Statistical Physics*, Pergamon Press, New York (1965).
[28] R. O. Zaĭtsev, Zh. Éksp. Teor. Fiz **50**, 1055 (1966) [Sov. Phys. JETP **23**, 702 (1966)].
[29] A. V. Svidzinskiĭ, *Spatially Inhomogeneous Problems of Superconductivity Theory* [in Russian], Nauka, Moscow (1982).
[30] B. Davison, *Theory of Neutron Transport*, Butterworths, London (1956).
[31] Yu. A. Izyumov, M. I. Katsnel'son, and Yu. N. Skryabin, *The Magnetism of Collectivized Electrons* [in Russian], Nauka, Moscow (1994).
[32] A. V. Zaĭtsev, Zh. Éksp. Teor. Fiz. **86**, 1742 (1984) [Sov. Phys. JETP **59**, 1015 (1984)].
[33] J. J. Hauser, H. C. Theuerer, and N. R. Werthamer, Phys. Rev. **142**, 118 (1966).
[34] M. Yu. Kupriyanov and V. F. Lukichev, Zh. Éksp. Teor. Fiz. **94**, No. 6, 139 (1988) [Sov. Phys. JETP **67**, 1163 (1988)].
[35] P. G. de Gennes, Rev. Mod. Phys. **36**, 225 (1964).

Translated by Eugene Yankovsky

Influence of unstable valence of cerium ions on the crystal field in ReNi compounds

V. N. Lazukov, P. A. Alekseev, E. S. Klement'ev, E. V. Nefedova, and I. P. Sadikov

Kurchatov Institute, 123182 Moscow, Russia

J.-M. Mignot

LLB, CEA-CNRS, CE-Saclay, 91191, Gif sur Yvette, France

N. B. Kol'chugina and O. D. Chistyakov

A.A. Baĭkov Institute of Metallurgy, 117911 Moscow, Russia

(Submitted 10 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1731–1747 (May 1998)

To examine the effect of hybridization of $4f$ electrons with conduction electrons on the crystal field potential using neutron spectroscopy, we studied the effects of the crystal electric field (CEF) in intermetallic compounds of the type ReNi, in which chemical substitution is followed by a transition of the cerium ions from an intermediate valence state to the Kondo state. Measurements were performed both on cerium ions in the compounds $Ce_{1-x}La_xNi$ ($x=0.5$, 0.8), where they have a whole-number population of the $4f$ shell, and on the paramagnetic impurity ion Nd in the series of compounds $Re_{1-x}Nd_xNi$ (Re=Ce, La, Y), in which the cerium ions are found either in an intermediate valence state or in the Kondo state. From the neutron inelastic magnetic scattering spectra on Nd ions, we have reconstructed the crystal field parameters in ReNi compounds and calculated the CEF level diagram of Ce ions in these compounds as functions of the inter-ion distances Re–Ni. The results of our calculations are in good agreement with the experimentally determined splitting diagram of the ground-state multiplet of the Ce ions. We have determined that as the degree of hybridization with the conduction electrons grows the CEF potential varies considerably and the effective splitting of the $4f$ shell of the cerium ions increases. The estimated energy scale of the splitting of the ground-state multiplet of the Ce^{3+} ions in the ReNi CEF ($\Delta_{CEF}\sim 15$ meV) turns out to be commensurate with the Kondo temperature ($T_K\sim 140$ K for CeNi. Analysis indicates that the CEF potential has a substantial effect on the formation of the valence-unstable ground state of the f shell in this compound. © *1998 American Institute of Physics.*
[S1063-7761(98)01205-0]

1. INTRODUCTION

In compounds based on rare-earth elements the interaction of the localized $4f$ shell with the crystal electric field (CEF) potential plays an extraordinarily important role in the formation of the ground state and the excitation spectrum of the $4f$ electrons, these being the two most important factors determining the physical properties of these compounds at low temperatures. At the same time, for compounds in which the rare-earth ions have an unstable $4f$ shell, in theoretical treatments the interaction with the CEF is typically considered to be insignificant in comparison with hybridization with the conduction electrons and, as a rule, is neglected.

The excitation spectrum of the $4f$ electrons of cerium ions in unstable-valence compounds obtained at temperatures $T>T_K$ (where T_K is the Kondo temperature) on polycrystalline samples using inelastic neutron scattering is usually a broad quasielastic distribution, which can be described with the help of the Lorentzian function. When the temperature is lowered, there also appears an inelastic component with a maximum at an energy on the order of a few tens of meV (see, for example, Ref. 1). Recently, a number of cerium-based unstable-valence compounds have been discovered, so-called Kondo insulators (CeNiSn, Ref. 2; $Ce_3Bi_4Pt_3$, Ref. 3), in which, with the help of neutron scattering a gap (pseudogap) was discovered in the spectrum of magnetic excitations of the $4f$ electrons along with the presence of a narrow gap in the density of electron states near the Fermi level. Moreover, a study on a CeNiSn single crystal showed that in CeNiSn the excitation spectrum of the $4f$ electrons has the following structure: two peaks with energies ~ 2 meV and ~ 4 meV. A gap in the excitation spectrum of the $4f$ electrons can be obtained theoretically as the result of hybridization of the degenerate ground-state multiplet of the f electron with states in the conduction band without taking other interactions into account. The spectrum above the gap in this case turns out to be broad and free of noticeable structure.[4,5] A possible reason for the appearance of features in the energy spectrum may have to do with the interaction of spin-liquid Fermi-type excitations with single-particle excitation of the CEF. A model was proposed in Ref. 6 which takes this interaction into account in order to explain the formation of the pseudogap in the excitation spectrum of $4f$ electrons of Ce ions in CeNiSn. A necessary condition of

this model is proximity of the excitation energies of the $4f$ electrons in the CEF to the Kondo temperature. In this case hybridization of the Kondo ground state and the CEF level can lead to a gap (pseudogap) in the spectrum of spin states. An experimental estimate of the CEF level diagram of the $4f$ electrons in CeNiSn has confirmed[7] that for this compound the energy of the first excited level and the half-width of quasielastic scattering characteristic of the Kondo interaction are close in value and have the scale ~ 4 meV.

A $4f$ gaplike excitation spectrum, similar to the spectrum of $Ce_3Bi_4Pt_3$, was discovered recently for a polycrystalline sample in the metallic compound CeNi at $T=12$ K (Ref. 8). The thermodynamic and kinetic properties of this compound are characteristic of systems with intermediate valence. The temperature dependence of the magnetic susceptibility has a maximum at $T=140$ K, and when the temperature is lowered further the compound behaves like a Pauli paramagnet with $\chi(0)\approx 2\times 10^{-3}$ emu/mol (Ref. 9). The coefficient of the electronic component of the specific heat γ is equal to 85 mJ/(mol·K^2) (Ref. 10). The temperature dependence of the magnetic component of the resistance (ρ_m) has a broad maximum at $T\approx 150$ K, and at low temperatures $\rho_m \sim T^2$, which is characteristic of a Fermi liquid.[9]

On the basis of these data the Kondo temperature was estimated as $T_K \approx 140$ K, which is close to the characteristic energies for other compounds with intermediate valence. At the same time, in contrast to the majority of compounds with intermediate valence that have been examined, the magnetic excitation spectrum in CeNi at $T<T_K$ has a gaplike form and cannot be fitted by a Lorentzian at low energies. The modified shape of the magnetic excitation spectrum in CeNi may be associated with structure similar to that found in CeNiSn single-crystal studies although T_K is considerable larger (by severalfold) for CeNi than for CeNiSn. With the aim of examining the possible role of the CEF potential in the formation of the ground state of unstable-valence ions it is of interest to estimate the scale of the interaction of the $4f$ electrons with the CEF in this compound.

2. EXPERIMENTAL APPROACH

For CeNi so far there have been no estimates of the interaction energies with the CEF. This situation is in part due to the fact that cerium ions in CeNi are found in an intermediate valence state and it is not possible to directly observe splitting of f-electron levels in the CEF. The valence of the cerium ions in CeNi ($v=3.15$ at $T=77$ K) turns out to depend on external parameters: it decreases when the temperature is increased[11] and grows when pressure is applied.[12] In connection with the possibility of acting on the valence of cerium ions, an estimate of the splitting diagrams of the $4f$-electron levels of Ce ions in a CEF can be approached in two ways.

The first way is connected with the application of ''negative'' chemical pressure upon substitution of some of the Ce ions by La. In this case, when the lanthanum concentration is increased to $x\sim 0.5$ in $Ce_{1-x}La_xNi$ compounds the cerium ions transition from the intermediate valence state to the Kondo regime, i.e., the degree of hybridization is decreased, and the valence becomes almost integer-valued. The transition of the Ce ions into the Kondo regime leads to the appearance in the neutron inelastic magnetic scattering spectrum of peaks associated with transitions between the $4f$-electron states due to splitting of the ground-state J-multiplet in the CEF.[8] Thus, the possibility opens up of indirectly examining CEF effects on cerium ions in compounds of the type $Ce_{1-x}La_xNi$. Since the immediate environment of the rare-earth ions consists of seven nickel ions (orthorhombic crystalline lattice of CrB type), when cerium is substituted by lanthanum no changes in the type of atoms in the immediate environment take place, only the distances to the neighboring nickel ions $R(\mathrm{Re}-\mathrm{Ni})$, and to first order it can be assumed that the ionic component of the CEF potential remains unchanged. Extrapolation of the values of the splitting energies of the CEF levels for cerium ions, which have a stable shell, with changes in the lattice parameters taken into account, allows one to estimate the scale of the splitting of the $4f$ shell in a CEF for compounds existing in an intermediate valence state.

To determine the diagram of the CEF levels of the cerium ions, when they are found in a state with a localized $4f$ shell, we can use the results of Ref. 8 for $x=0.8$ (0–7–15 meV). For $Ce_{0.5}La_{0.5}Ni$, in which Ce is found near a transition between the Kondo regime and an intermediate valence state, low-temperature measurements are insufficient: the widths of the transitions between the CEF levels, determined by the degree of hybridization of the f electrons with the conduction electrons, are comparable to their energies at $T=12$ K and it is impossible to unambiguously reconstruct the level diagram. Since the valence of the cerium ions in CeNi decreases with increase of the temperature, it may be expected that for the sample with $x=0.5$, with increase of the temperature the degree of hybridization will also decrease. This should lead to a decrease of the widths of the transitions between the CEF levels in the neutron spectra, i.e., measurements of inelastic neutron scattering spectra over a wide temperature range can make it possible to obtain the diagram of CEF levels of the cerium ions in this compound.

The second way of estimating the splitting diagram of the $4f$ shell is to determine the parameters of the CEF potential with the aid of measurements of CEF effects on inelastic magnetic scattering by an impurity paramagnetic ion (paramagnetic tag) in ReNi-based matrices and is similar to the approach used in Ref. 7. On the basis of the experimental data, it is possible to reconstruct the parameters of the CEF potential in ReNi compounds and calculate the level diagram of the cerium ions in CeNi. In addition, from a comparison of the data for neodymium ions in compounds containing cerium ions and without them, it is possible to determine the influence of the unstable valence of the cerium ions on the CEF potential. In principle, these methods can give varying results for the CEF potential at the site of the rare-earth ion since hybridization can lead to an additional ''contribution'' to the effective ''CEF splitting'' observed for the Ce ion in an unstable f-shell.

In the present work we have implemented both of the above-indicated methods and have compared the results.

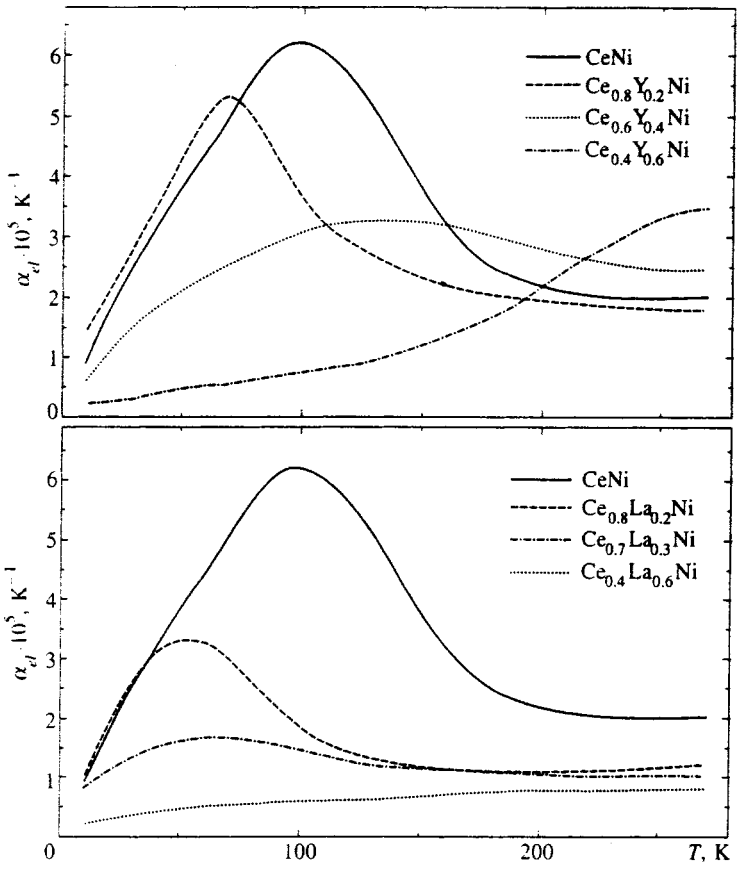

FIG. 1. Temperature dependence of the electronic component of the thermal expansion coefficient for the compounds $Ce_{1-x}Y_xNi$ and $Ce_{1-x}La_xNi$.

3. MEASUREMENT TECHNIQUE AND SAMPLES

To perform the measurements, it is necessary first of all to choose the composition of the samples in which the cerium ions are found in a state with partially delocalized $4f$ shells. The state of the cerium ions with intermediate valence in the investigated samples was determined on the basis of the experimentally obtained dependence of the valence state on the interion distances. To this end, we performed precision x-ray measurements of the lattice parameters of $Ce_{1-x}(La,Y)_xNi$ substituted compounds over a wide temperature range (10–300 K). The use of La and Y ions leads to an increase (La) or decrease (Y) of the lattice parameters. Here the absence of an f shell in these ions allows one to avoid additional contributions to the temperature variations of the lattice parameters from CEF effects. On the basis of these measurements we calculated the distances between the rare-earth ions and the neighboring nickel ions R(Re–Ni). Use of the distance R(Re–Ni) as a physical parameter makes it possible for compounds with low crystal symmetry to represent the influence of changes in the lattice constants on only one parameter. From the temperature dependence of the R(Re–Ni) distances obtained from x-ray diffraction measurements we determined the thermal expansion coefficients for the $Ce_{1-x}(La,Y)_xNi$ samples and isolated the electronic component of this coefficient ($\alpha_{el}=\alpha_{ReNi}-\alpha_{LaNi}$) (Fig. 1). For all the yttrium-bearing compounds and for some with lanthanum a maximum is observed in $\alpha_{el}(T)$. A maximum in the temperature dependence of the electronic component of the thermal expansion coefficient of intermediate-valence compounds is usually associated with a change in the mean number of localized electrons in the $4f$ shell of the cerium ions with temperature.[13] Substitution of Ce by La increases the R(Ce–Ni) distance (Fig. 2) and for $x>0.3$ the lack of a maximum is an indication of a transition to a stable state of

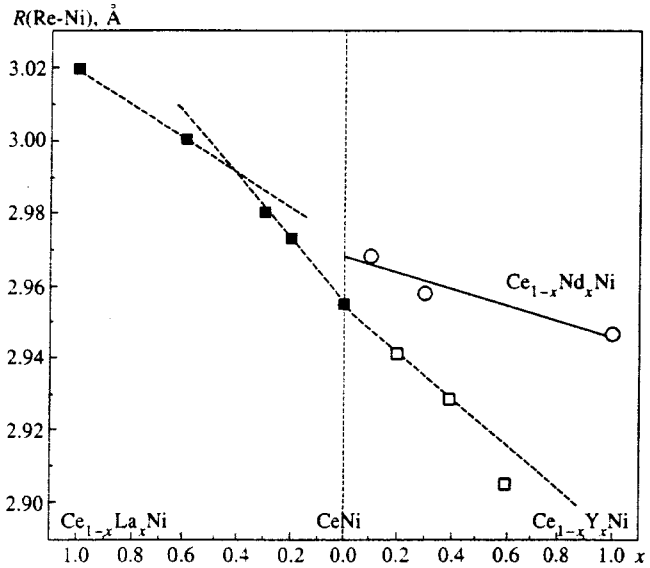

FIG. 2. Concentration dependence of the distances R(Re–Ni) for the compounds $Ce_{1-x}La_xNi$ (■), $Ce_{1-x}Y_xNi$ (□), and $Ce_{1-x}Nd_xNi$ (○) at room temperature. The measurement error is less than the width of the symbols. The dashed lines were drawn by eye. The solid line is Végard's law for the compounds $Ce_{1-x}Nd_xNi$ for $x>0$.

the f electron. This conclusion is also supported by the change in the slope in the concentration dependence of $R(\mathrm{Re}-\mathrm{Ce})$ at $x\sim 0.5$ (Fig. 2). An increase in the Y ion concentration, on the other hand, decreases the $R(\mathrm{Ce}-\mathrm{Ni})$ distance and increases the valence.[14] It turns out that for $R(\mathrm{Ce}-\mathrm{Ni})>2.98$ Å the cerium ions have an integer-valued population of their $4f$ shell over the entire investigated temperature interval. For CeNi compounds the position and magnitude of the maximum in the electronic component fall out of the observed systematic dependence on x: its position turns up at a higher temperature and with a larger value of the amplitude α_{el} than is the case even for the sample $Ce_{0.8}Y_{0.2}Ni$ (Fig. 1a), which has smaller values of $R(\mathrm{Re}-\mathrm{Ni})$. This circumstance is possibly connected with the appearance of an additional contribution to the thermal expansion coefficient at low temperatures from the Ce–Ce interaction in the intermediate-valence lattice upon formation of the ground state of the cerium ions. This interaction can also affect the interion distances in the regular (in the Ce ions) rare-earth sublattice and is suppressed by the disordering attendant to substitution.

On the basis of the data on the dependence of the valence state on the interion distance $R(\mathrm{Re}-\mathrm{Ni})$ we determined the composition of the samples to determine the parameters of the CEF potential and prepared polycrystalline samples of $Nd_xCe_{1-x}Ni$ ($x=0.1,\ 0.3$), in which the Ce ions are found in an intermediate valence state, and $Nd_{0.1}La_{0.5}Ce_{0.4}Ni$, in which the cerium ions are found in the Kondo regime. To determine the effect of variations of the lattice parameters on the CEF potential at the rare-earth ion sites, we used samples of $Nd_{0.1}La_{0.5}Y_{0.4}Ni$ and $Nd_{0.3}La_{0.7}Ni$, which do not contain cerium ions. The same polycrystalline samples of $Ce_{1-x}La_x^{60}Ni$ ($x=0.5,\ 0.8$) were used as in Ref. 8. All polycrystalline samples were prepared in an arc furnace in an argon atmosphere. X-ray diffraction analysis of the samples showed that the prepared compounds have orthorhombic structure of the type CrB. The fraction of other phases present was less than 3%. The mass of the polycrystalline samples used in the neutron scattering measurements varied depending on the cerium concentration from 60 to 120 g.

Inelastic neutron scattering experiments on samples containing Nd ions were performed on 4F2 and 2T three-axis spectrometers outfitted with cold (4F2) and hot (2T) neutron sources on the Orphée neutron reactor (LLB, Saclay), and also on the MIBEMOL time-of-flight spectrometer (LLB, Saclay).

The measurements on the three-axis spectrometers were performed using the $Q=\mathrm{const}$ method with $E_f=8.05$ meV on the 4F2 and $E_f=14.7$ meV on the 2T instrument. As the monochromator and analyzer we used pyrographite (the (002) reflection), and with the help of the pyrographite filter we suppressed the higher orders of reflection. The FWHM energy resolution of the elastic peak was $\delta E=0.44$ meV (4F2) and $\delta E=0.9$ meV (2T). We studied energy transfer in the range 0–14 meV at temperatures 1.8–160 K (4F) and 0–30 meV at 12–80 K (2T). The values of the neutron momentum transfer varied from 0.9 to 5 $\mathrm{Å}^{-1}$. For the $Nd_{0.1}La_{0.5}Ce_{0.4}Ni$ sample the position of the level with $E\cong 15$ meV was determined from measurements on the KDSOG-M time-of-flight spectrometer with reverse geometry (IBR-2, JINR, Dubna) at $T=10$ K.

The low-energy part of the inelastic neutron scattering spectrum for the $Nd_{0.3}La_{0.5}Ce_{0.4}Ni$ sample was investigated on the MIBEMOL time-of-flight spectrometer with fixed energy of the incident neutrons $E_i=3.27$ meV at 15 and 45 K. The energy resolution at $E=0$ was 0.16 meV.

The inelastic neutron scattering spectrum was separated into its magnetic and phonon components on the basis of differences in the dependence of the intensity of magnetic and nuclear scattering on the magnitude of the momentum transfer Q. For the magnetic component the dependence of the intensity on Q was modeled by the form-factor for Nd^{3+} ions,[15] and for the phonon component we used the quadratic law $S(Q)\sim Q^2$. We analyzed the inelastic neutron scattering spectra corresponding to small ($0.9-2$ $\mathrm{Å}^{-1}$) and large ($4-6$ $\mathrm{Å}^{-1}$) values of momentum transfer. The correctness of the procedure was confirmed by coincidence within the limits of experimental error of the calculated and measured spectra for intermediate values of Q. The procedure is described in detail in Ref. 7. Figure 3 presents the inelastic neutron scattering spectrum for $Nd_{0.3}La_{0.7}Ni$, measured on the 4F2 spectrometer with $Q=1.76$ $\mathrm{Å}^{-1}$, with the magnetic and phonon contributions plotted separately. It was found that at small Q and low temperatures ($T<60$ K) the phonon fraction did not exceed 10%.

To determine the CEF level diagram of the cerium ions when they are found in the state with the localized $4f$ shell, we used the results of measurements on samples of $Ce_{1-x}La_xNi$ ($x=0.5,\ 0.8$). Measurements of the inelastic neutron scattering spectra for $x=0.5$ at 12, 50, 100, and 150 K were performed on the HET time-of-flight spectrometer (ISIS, RAL) with incident neutron energy $E_i=80$ meV (Ref. 16). The technique for separating out the magnetic component is analogous to the procedure described in Ref. 8. Experimental data for the $Ce_{0.2}La_{0.8}Ni$ sample are given in Ref. 8.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Parameters of the CEF potential in $Nd_xRe_{1-x}Ni$

The CEF potential at the neodymium ion sites in orthorhombic compounds of the type ReNi with point group symmetry C_{2v} can be described within the framework of a phenomenological approach with the help of a Hamiltonian with nine independent parameters. To determine the parameters of the CEF Hamiltonian we used the parametrization suggested in Ref. 17:

$$H_{cf}=W\sum_{nm} x_{nm}\tilde{O}_n^m,$$

where $\tilde{O}_n^m=\hat{O}_n^m/F_{nm}$, $\hat{O}_n^m$ is the Stevens operator,[18] and the normalized factors F_{nm} are tabulated in Ref. 17. The parameter W is a scale factor. The CEF parameters x_{nm} are related by the condition $\Sigma|x_{nm}|=1$.

In the case under consideration, to determine the parameters of the CEF potential it is necessary to obtain experimental data containing at least nine independent parameters

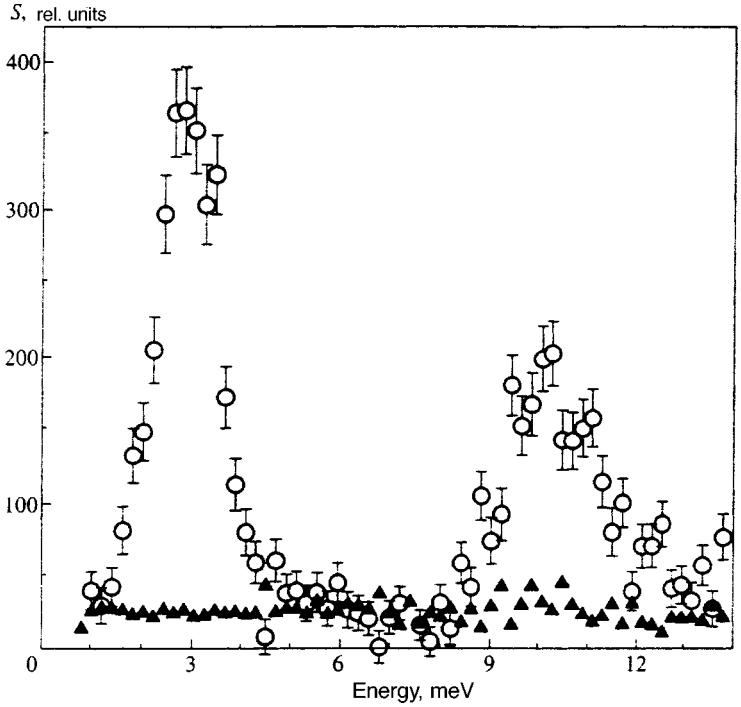

FIG. 3. Magnetic (circles) and phonon (triangles) components of the neutron inelastic scattering spectra for the sample $Nd_{0.3}La_{0.7}Ni$. The measurements were made at $T=5.8$ K on a 4F2 spectrometer by the $Q=$const method for $Q=1.76$ Å^{-1}.

such as the energy positions of the peaks and their intensities. The Nd^{3+} ions have total angular momentum $J=9/2$, i.e., the ground-state multiplet has degeneracy equal to $2J+1=10$. The interaction of the f shell with the CEF can lower the degeneracy to two. Thus, the ground-state J multiplet of the $4f$ shell in NdNi splits into five levels (doublets). Measurement of the inelastic neutron scattering spectra at low temperatures allows one to measure transitions only from the ground state to all excited levels of the ground-state multiplet and thereby indirectly obtain the splitting diagram of the $4f$ level for Nd^{3+}. The diagrams of the Nd ion $4f$ levels so obtained are depicted in Fig. 4 for all the investigated samples as functions of the distance R(Re–Ni) between the rare-earth ion and the neighboring Ni ions. The figure also plots results of a calculation of the energy levels of the $4f$ electrons of the Nd^{3+} ions for two of the distances for which experimental data exist. The CEF parameters and energy levels were calculated from relations of the CEF superposition model[19] which allows for the dependence of the geometrical coordination factors on the lattice parameters. As our starting parameters we took the CEF parameters obtained from experimental data for the $Nd_{0.3}La_{0.7}Ni$ sample (diagram 5 in Fig. 4). For changes in the interion distances on the order of 50% of the maximum change in our series of experiments the calculated level diagram (squares in Fig. 4) is found to be in good agreement with a linear interpolation (solid line in Fig. 3) between the experimentally obtained level diagrams. With further variation of the interion distances the calculated positions of the energy levels begin to differ substantially from their experimental values. In view of this, as a dependence linking the CEF parameters and the interion distances, we used a linear interpolation of the experimental results for the $Nd_x(La,Y)_{1-x}Ni$ samples.

From the results presented in Fig. 4 it follows that as the interion distance R(Re–Ni) decreases, the splitting of the $4f$ shell increases for Nd^{3+} ions in compounds not containing cerium ions (the ''normal'' matrix, corresponding to diagrams 1 and 5 in Fig. 4). In compounds containing cerium ions, the observed changes in the energy levels follow the opposite trend. In $Nd_{0.1}Ce_{0.4}La_{0.5}Ni$, in which cerium has a localized f shell (the Kondo regime) (diagram 4 in Fig. 4), only an insignificant decrease of the magnitude of splitting is observed. At the same time, for compounds containing ce-

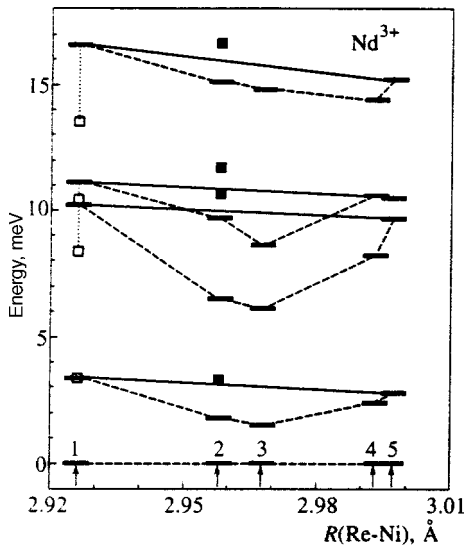

FIG. 4. Experimentally determined splitting diagrams for the $4f$ levels of Nd^{3+} in ReNi compounds as functions of the distance between the rare-earth ion and the neighboring Ni ions [R(Re–Ni)] (1—$Nd_{0.1}La_{0.5}Y_{0.4}Ni$, 2—$Nd_{0.3}Ce_{0.7}Ni$, 3—$Nd_{0.1}Ce_{0.9}Ni$, 4—$Nd_{0.1}Ce_{0.4}La_{0.5}Ni$, 5—$Nd_{0.3}La_{0.7}Ni$). The solid line corresponds to measurement of the positions of the levels of the $4f$ electrons as functions of the distance R(Re–Ni) and was obtained by interpolation. The symbols ■ and □ correspond to the calculated results (see text). The dashed lines were drawn by eye.

rium ions with intermediate valence (diagrams 2 and 3 in Fig. 4) the decrease in the energy levels is much greater.

Attention is drawn to the nonlinearity of the dependence of R(Re–Ni) for $Ce_{1-x}Nd_xNi$ on x in the region of small x (Fig. 2). Substitution of 10% Ce by Nd leads to an increase in R(Re–Ni) in comparison with CeNi. At the same time, with further growth of x from 0.1 to 0.3, as could be expected from a comparison of the neighboring distances R(Nd–Ni)=2.947 and R(Ce–Ni)=2.955 Å, the distance R(Re–Ni) decreases with growth of x, i.e., the CeNi lattice has noticeably smaller parameters than could be expected from the extrapolation from $x=0$ of the values of R(Re–Ni) for samples containing neodymium ions. The anomaly in the concentration dependence of the distance to the neighboring Ni ions, observed in pure CeNi, is in accord with the ''nonmonotonicity'' of the data on the electronic component of the thermal expansion coefficient (Fig. 1): apparently, in the ordered rare-earth sublattice an additional interaction is present, which is probably the reason for the observed effects. The presence of an additional interaction is apparently also the reason for the stronger deviation of the level diagram for the sample with $x=0.1$ from the interpolational dependence in comparison with the results for $x=0.3$ (diagrams 3 and 2 in Fig. 4, respectively).

The low-temperature measurements of the inelastic neutron scattering spectra (i.e., when only the ground state is populated) yield eight independent parameters: the energies of the four excited levels and the four values of the deviations of the intensities for the transitions from the ground state. Additional parameters can be obtained by measuring the intensities of the transitions between the excited levels of the $4f$ electrons, which become increasingly populated as the temperature is increased. Thus, for example, the measurements of the inelastic neutron scattering spectra of the $Nd_{0.3}La_{0.7}Ni$ sample in the temperature interval 5–160 K made it possible to determine the values of the nine independent experimental parameters: the energies of all the excited levels (Fig. 4) and the ratios of the intensities for the five transitions (see Table I for $Nd_{0.3}La_{0.7}Ni$) (The intensity of the transition from the ground state to the first excited state was used to calibrate the measurements performed on different spectrometers.) But for the samples with Nd ions in a cerium matrix ($Nd_{0.3}Ce_{0.7}Ni$ band $Nd_{0.1}Ce_{0.9}Ni$) we determined 10 independent experimental parameters: the energies of all the excited states (Fig. 4) and the ratios of intensities for six transitions (these data are given in Table I for $Nd_{0.3}Ce_{0.7}Ni$). Thus, for the samples with impurity ions in lanthanum and cerium matrices we obtained experimental information which turned out to sufficient to uniquely determine the parameters of the CEF Hamiltonian. First we reconstructed the CEF parameters for the compounds $Nd_{0.3}La_{0.7}Ni$, $Nd_{0.3}Ce_{0.7}Ni$, and $Nd_{0.1}Ce_{0.9}Ni$. The procedure for calculating the CEF parameters, based on a complete grid search of CEF parameter space, is described in detail in Refs. 20 and 21. For the remaining samples, the values of the CEF parameters were obtained by using the CEF parameters of $Nd_{0.3}La_{0.7}Ni$ as the starting parameter set in the local minimization procedure.[20] A grid search of the space of all possible values of the CEF parameters showed that there exists a unique parameter set which is adequate to the available experimental information for each of the investigated samples. Final results for the parameters of the CEF potential in Walter's parametrization[17] are listed in Table II for all the investigated samples.

TABLE I. Experimentally determined values of the ratios (with relative accuracy roughly 30%) of the squares of magnitudes of the matrix elements of the transitions between the CEF levels $|i\rangle$ and $|j\rangle$ to the probability of the transitions between the states $|1\rangle$ and $|2\rangle$ ($|\langle j|J_\perp|i\rangle|^2/|\langle 2|J_\perp|1\rangle|^2$) for $Nd_{0.3}La_{0.7}Ni$ and $Nd_{0.3}Ce_{0.7}Ni$.

	$Nd_{0.3}La_{0.7}Ni$						$Nd_{0.3}Ce_{0.7}Ni$				
i/j	1	2	3	4	5	i/j	1	2	3	4	5
1	4.6	1	0.61	0.30	1.1	1	-	1	0.1	0.58	0.38
2		-	0.85	<0.1	0.44	2		-	<0.15	<0.15	0.36
3			-	<0.2	<0.2	3			-	0.62	0.70
4				-	<0.4	4				-	1.02
5					-	5					-

4.2. Ce level diagram in $Ce_{1-x}La_xNi$

Figure 5 plots the results of measurements of the temperature dependence of the magnetic component of the inelastic neutron scattering on the $Ce_{0.5}La_{0.5}Ni$ sample. In contrast to the measurements performed at $T=12$ K (Ref. 8), two peaks are observed in the neutron spectra as the tempera-

TABLE II. Values of the nearest-neighbor distances between the rare-earth ions and Ni [R(Re–Ni), $\Delta R \cong \pm 0.001$ Å] and the CEF parameters in Walter's parametrization[17] for the investigated samples: $W(\Delta W \simeq \pm 0.02$ meV) and $x_{nm}(\Delta x_{nm} \simeq \pm 0.015)$.

	$Nd_{0.3}La_{0.7}$–Ni	$Nd_{0.1}Ce_{0.9}Ni$	$Nd_{0.3}Ce_{0.7}Ni$	$Nd_{0.1}La_{0.5}Y_{0.4}Ni$	$Nd_{0.1}Ce_{0.4}La_{0.5}Ni$
R(Re–Ni), Å	2.997	2.968	2.958	2.926	2.992
W, meV	0.49	0.56	0.60	0.52	0.46
x_{20}	0.225	0.165	0.152	0.191	0.226
x_{22}	0.066	0.199	0.205	0.029	0.056
x_{40}	0.020	−0.101	−0.091	0.013	0.013
x_{42}	−0.049	−0.020	−0.029	−0.060	−0.042
x_{44}	0.069	0.030	0.048	0.067	0.021
x_{60}	−0.394	−0.292	−0.276	−0.356	−0.400
x_{62}	−0.043	−0.040	−0.048	−0.016	−0.061
x_{64}	−0.057	−0.081	−0.087	−0.061	−0.098
x_{66}	0.080	−0.072	−0.064	0.228	0.072

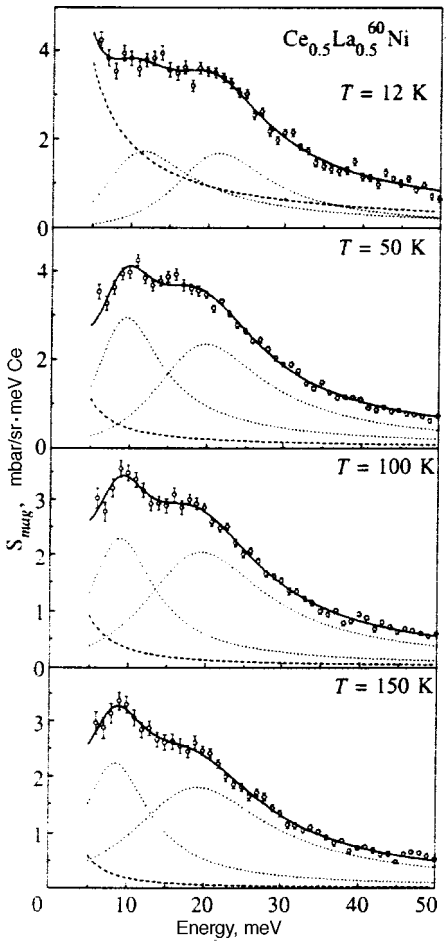

FIG. 5. Magnetic component of neutron inelastic scattering in the sample $Ce_{0.5}La_{0.5}^{60}Ni$, measured in the temperature interval 12–150 K. The solid lines are fits to the spectra using two inelastic (dotted lines) and one quasi-elastic (dashed line) Lorentzian.

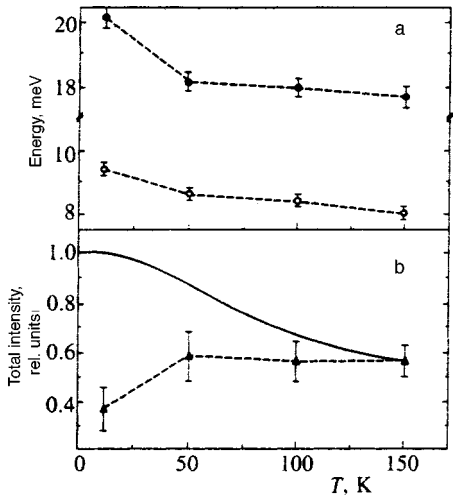

FIG. 6. a) Temperature dependence of the energy of the inelastic peaks in the spectra of the magnetic component of inelastic neutron scattering from a sample of $Ce_{0.5}La_{0.5}^{60}Ni$. (Lines drawn by eye.) b) Temperature dependence of the relative total intensity of the inelastic peak with $E \approx 18$ meV; solid line—calculation of the variation of the population of the ground state based on the level diagram of the cerium ion; triangles—experimental data (dashed line drawn by eye).

ture is raised, more distinctly and with larger widths than in $Ce_{0.2}La_{0.8}Ni$. These peaks may be associated with transitions between CEF levels. Fitting the spectra with the help of two inelastic and one elastic Lorentzian gives the following energies: $E \approx 9$ meV ($\Gamma/2 \sim 5$ meV) and 18 meV ($\Gamma/2 \sim 10$ meV). At low temperatures, when only the ground state is populated, these peaks correspond to two transitions from the ground state, which allows one to directly determine the CEF level diagram of the $4f$ electrons of the cerium ions in this compound.

Note that raising the temperature should lead to population of the first excited state, i.e., it should be accompanied by a decrease in the intensity of the transitions from the ground state and an increase in the intensity of the transition from the first excited state ($E \approx 9$ meV). The energy of this transition is approximately equal to the energy of the transition from the ground state to the first excited state. In other words, the peak at $E \approx 18$ meV should decrease as the temperature is increased while the peak with $E \approx 9$ meV can vary arbitrarily since its intensity is determined by the ratio of the matrix elements of the two transitions contributing to this peak at intermediate temperatures. Figure 6b plots the temperature dependence of the intensity of the $E \approx 18$ meV peak and a calculation of the variation of the intensity in the single-ion approximation, due to temperature population of the ground state. For ease of comparison, the data are plotted in relative units and scaled to the values at $T = 150$ K. As can be seen from the figure, as the temperature is decreased the intensity of the observed experimental transition does not increase, as follows from calculation, but remains constant and for $T < 50$ K it even decreases. What is more, a noticeable increase in the transition energies takes place in this temperature interval (Fig. 6a). A similar change in the magnetic excitation spectrum of the cerium ions with decrease of the temperature was observed for the heavy-fermion compound $CeAl_3$ (Ref. 22) and was interpreted as a consequence of formation of the Kondo ground state as the temperature was lowered. In the compound $Ce_{0.5}La_{0.5}Ni$, in which the cerium ions are found in the Kondo regime, it appears that a new singlet ground state is also formed and the excitation spectrum of the $4f$ electrons is correspondingly transformed when the temperature is lowered.

As R(Re–Ni) decreases with increasing Ce concentration in (Ce,La)Ni compounds, the hybridization energy increases. A comparison of the obtained level diagrams for the $Ce_{0.5}La_{0.5}Ni$ samples (0–9–18 meV for $T \geqslant 50$ K) and $Ce_{0.2}La_{0.8}Ni$ samples (0–7–15 meV) (Ref. 8) shows that as the hybridization energy increases, an increase is observed in the splitting energy of the $4f$ levels combined with a broadening of the transitions between these levels. Note that an increase in the hybridization energy also leads to an increase in the temperature of the maximum in the thermal expansion coefficient (Fig. 1), which is consistent with ideas about the connection between the position of the maximum and the hybridization energy.[13]

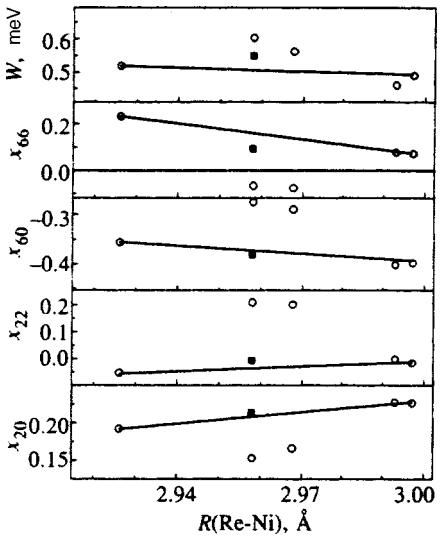

FIG. 7. Values of W and the CEF parameters x_{30}, a_{22}, x_{60}, and x_{66}, derived from the experimental data, plotted versus distance R(Re–Ni) for $Nd_{0.1}La_{0.5}Y_{0.4}Ni$, $Nd_{0.3}Ce_{0.7}Ni$, $Nd_{0.1}Ce_{0.9}Ni$, $Nd_{0.1}Ce_{0.4}La_{0.5}Ni$, and $Nd_{0.3}Na_{0.7}Ni$ (see Fig. 4). The squares (■) plot the results of calculation (see text).

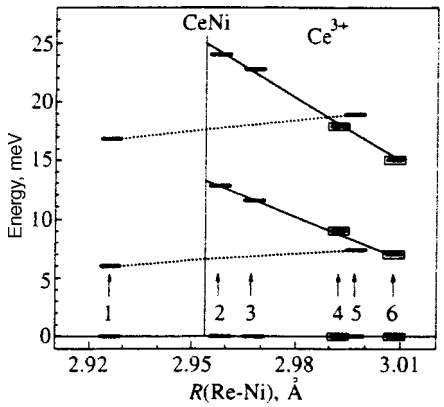

FIG. 8. Splitting diagrams of the $4f$ levels of the Ce^{3+} ions in a crystal electric field, derived from measurements of inelastic neutron scattering directly from the Ce ions ($Ce_{1-x}La_x^{60}Ni$: 4—$x=0.5$, 6—$x=0.8$) and calculated from values determined experimentally from measurements on the paramagnetic tag Nd of the CEF parameters in the ReNi compound (1—$Nd_{0.1}La_{0.5}Y_{0.4}Ni$, 2—$Nd_{0.3}Ce_{0.7}Ni$, 3—$Nd_{0.1}Ce_{0.9}Ni$, 5—$Nd_{0.3}Na_{0.7}Ni$) versus the distance between the rare-earth ion and the neighboring Ni ions [R(Re–Ni)]. The solid line is an extrapolation of the positions of the $4f$-electron levels of the Ce^{3+} ions plotted versus R(Re–Ni) according to data for samples 4 and 6. The dotted line is an interpolation of the position of the $4f$ level plotted versus R(Re–Ni) in a ''normal'' matrix. The solid vertical line shows the value of R(Re–Ni) for CeNi.

4.3. Influence of intermediate-valence state on level splitting and potential parameters

Figure 7 plots as functions of R(Re–Ni) some of the CEF potential parameters that have relatively large values and at the same time appear to be most sensitive to delocalization of the $4f$ electrons. All of the values of the CEF parameters obtained for compounds with cerium ions in the Kondo regime coincide well with the quantities obtained by interpolation of the CEF parameters obtained for a ''normal'' matrix (see Sec. 3.1). The only exception is the value of W, which is somewhat smaller. For the matrix with intermediate valence the deviation of all the plotted parameters is significant, which is evidence of an appreciable redistribution of charge brought about by partial delocalization of the $4f$ electrons of the cerium ions in the intermediate-valence regime. It is striking that, as was the case in the measurements on the other intermediate-valence compound $CeNi_5$ (Ref. 23), considerable variation was observed in the sixth-order parameters for the CEF potential at the paramagnetic tag, to which only electrons with orbital angular momentum $l \geqslant 3$ can contribute. In the compounds under study, only f electrons possess such large angular momentum, and these variations indicate either partial delocalization of these electrons or formation (as a result of hybridization of the f electrons with states in the conduction band) of a new electron density distribution with f-type symmetry from the electron wave functions (e.g., of d and p type) of lower symmetry.

The CEF Hamiltonian for the cerium ions, for which the total angular momentum $J=5/2$, contains five independent terms. To determine these parameters only on the basis of measurements of inelastic neutron scattering spectra on the Ce ions in CeNi, it is necessary to determine at least five independent quantities characterizing the splitting of the ground-state multiplet in the CEF (the energy levels, and the probabilities of the transitions between them). In principle, the required number of parameters of the CeNi spectrum can be obtained. However, in the given situation, where the transition widths are comparable to the energy levels but the energies of the transition from the ground state to the first excited state and between excited states are roughly equal, it is quite difficult to obtain the transition intensities directly with reasonable accuracy. Furthermore, sixth-order terms are absent from the CEF Hamiltonian for the cerium ions, which precludes a direct comparison of the CEF potential parameters obtained from measurements on Nd and Ce ions. With this in mind, to compare data on the CEF potential obtained by different methods, we calculated the splitting diagrams of the $4f$ levels of the cerium ions in CeNi from the CEF parameters obtained from measurements on neodymium ions in ReNi compounds. These results, and also results of direct measurements of the energy levels of the Ce ions, are plotted in Fig. 8. It is evident from a comparison of the calculated splitting diagrams that the dependence of the energies of the excited levels on the interion distances, obtained by two different methods, are in good agreement. On the basis of this fact, we can conclude that the CEF potential parameters determined on the basis of measurements with a paramagnetic tag (Nd) give a realistic estimate of the CEF in ReNi compounds.

As follows from our calculation of the CEF level diagram of the Ce ions, based on measurements with the paramagnetic tag Nd (Fig. 8), with decrease of R(Re–Ni), i.e., with increase of the degree of hybridization of the $4f$ electrons of the Ce ions in the ReNi matrix, the variation of the CEF potential is manifested in an increase of the CEF energy levels. The same trend is also seen in results obtained by direct measurement on Ce ions (Fig. 8), i.e., the two approaches to determining the CEF potential give identical results. At the same time, in the matrix not containing cerium

ions (sample 1), a decrease in the Re–Ni distance leads to an insignificant decrease in the calculated CEF energy levels for the cerium ions. It can be concluded that hybridization of the $4f$ electrons with conduction electrons in intermediate-valence compounds makes a substantial additional contribution to the energy splitting of the ground-state multiplet of the $4f$ shell of the cerium ions.

Comparison of the CEF potential parameters obtained earlier with the help of a Pr impurity tag[21] with the results of the present work demonstrates qualitative agreement of the obtained data: only an insignificant change in the scale factor W is noticed in the Kondo regime while in the intermediate-valence state along with a change in W a pronounced change is observed in a number of CEF parameters, above all the sixth-order parameters. Calculation of the Ce level diagram on the basis of the CEF parameters in ReNi compounds, obtained from measurements with the Pr tag, also leads to an increase in the splitting energies of the CEF levels of the Ce ions when going from the Kondo regime to the intermediate-valence regime. However, the values of the CEF energy levels of the cerium ions turn out to be approximately half those obtained in the present work. The marked difference in the CEF parameters reconstructed from the inelastic neutron scattering spectra on Pr and Nd ions is also the case for other nickel compounds, in particular for $ReNi_5$ (Refs. 24 and 25). It is possible that the difference in the results obtained on Pr and Nd ions has to do with the pronounced perturbation of the electronic subsystem caused by the Pr ions, which can be near to a valence-unstable state. Therefore the use of Pr ions as the paramagnetic tag does not always make it possible to obtain quantitative information about the CEF potential.

Extrapolation of the excited energy levels of the Ce ions as functions of the Re–Ni distance makes it possible to estimate the possible splitting diagram of the cerium ions in a CeNi matrix in the case when the cerium ions have a localized $4f$ shell: 0, $\sim$14, $\sim$25 meV. On the basis of specific heat and magnetic susceptibility measurements in CeNi, we arrive at the estimate $T_K \sim 120-150$ K.[9,10] If we use an empirical relation between the maximum in the magnetic component of the inelastic neutron scattering spectrum ($E \approx 45$ meV, Ref. 8) and the Kondo temperature determined from thermodynamic characteristics,[26,27] then we obtain a value close to $T_K \approx 130$ K. The quantity T_K turns out to be approximately equal to the energy of the first excited CEF level. Thus, as in the case of CeNiSn, the energy scale of the interaction of the $4f$ electrons with the CEF in CeNi turns out to be similar to the energy of hybridization with the conduction electrons, and neither of these interactions can be excluded from consideration in an analysis of the formation of the ground state of the Ce ion, i.e, the CEF potential probably plays a role in the formation of the ground state of the f shell in CeNi. It may be surmised that it is precisely this closeness between the hybridization energy and the magnitude of the splitting of the $4f$ level in a CEF that leads to novel magnetic excitation spectra.

5. CONCLUSION

The studies reported here of the CEF potential in ReNi compounds make it possible to estimate the possible magnitude of the splitting of the $4f$ shell of cerium ions in the intermediate-valence compound CeNi, which turns out to be comparable to the Kondo temperature. The CEF level diagram of the $4f$ electrons of the Ce ions was obtained both by extrapolation of inelastic neutron scattering data on cerium ions with almost integer-valued population of the f shell in compounds near the transition to the intermediate-valence state and from a calculation of the CEF parameters reconstructed from experimental data for Nd impurity ions in CeNi-based valence-unstable compounds. The good agreement of the results confirms the applicability of the paramagnetic tag method for obtaining a realistic estimate of the magnitude of the CEF potential in intermediate-valence compounds. Note that as in the case of $CeNi_5$ (Ref. 23), a substantial change in the sixth-order CEF parameters, which are determined mainly by electrons with wave-function symmetry corresponding to f electrons, is observed in CeNi.

ACKNOWLEDGMENTS

We are indebted to A. S. Mishchenko, K. A. Kikoin, and Yu. M. Kagan for their support and useful discussions, and E. A. Goremychkin and A. Yu. Muzychka for assistance with the measurements. P. A. Alekseev and E. S. Klement'ev thank the workers of the Leon Brillouin Laboratory (LLB, France) for their cooperation in carrying out the neutron measurements, V.N.L. and I.P.S. are grateful to the administration of Rutherford Laboratory (RAL, England) for the opportunity to carry out experiments on the HET spectrometer. This work was carried out with the financial support of the Russian Fund for Fundamental Research (Project No. 95-02-04734) and the State Scientific and Technical Program ''Current Topics in the Physics of Condensed Media.''

[1] M. Loewenhaupt and K. H. Fischer, *Handbook on the Physics and Chemistry of Rare Earths* **16**, 2 (1993).
[2] T. E. Mason, G. Aeppli, A. P. Ramirez *et al.*, Phys. Rev. Lett. **69**, 490 (1992).
[3] A. Severing, J. D. Thompson, P. C. Canfield *et al.*, Phys. Rev. B **44(13)**, 6832 (1991).
[4] A. J. Fedro and S. K. Sinha, in *Valence Fluctuations in Solids*, I. M. Falikov, W. Hanke, and M. B. Maple (eds.), North-Holland, Santa Barbara (1981).
[5] D. L. Huber, Phys. Rev. B **28(2)**, 860 (1983).
[6] Yu. Kagan, K. A. Kikoin, and A. S. Mishchenko, Phys. Rev. B **55(18)**, 12 348 (1997).
[7] P. A. Alekseev, E. S. Klement'ev, V. N. Lazukov *et al.*, Zh. Éksp. Teor. Fiz. **106**, 1228 (1994) [JETP **79**, 665 (1994)].
[8] V. N. Lazukov, P. A. Alekseev, E. S. Clementyev *et al.*, Europhys. Lett. **33(2)**, 141 (1996).
[9] D. Gignoux, F. Givord, and R. Lemaire, J. Less-Common Met. **94**, 165 (1983).
[10] Y. Isikawa, K. Mori, T. Mizushima *et al.*, J. Magn. Magn. Mater. **70**, 385 (1987).
[11] V. A. Shaburov, Yu. P. Smirnov, A. E. Sovestnov *et al.*, JETP Lett. **41**, 259 (1985).
[12] D. Gignoux, C. Vettier, and J. Voiron, J. Magn. Magn. Mater. **70**, 388 (1987).
[13] E. Muller-Hartmann, Solid State Commun. **31**, 113 (1979).
[14] P. A. Alekseev, E. S. Clementyev, V. N. Lazukov *et al.*, Phys. Met. Metallogr. **77(6)**, 611 (1994).
[15] M. Blume, A. J. Freeman, and R. E. Watson, J. Chem. Phys. **37**, 1245 (1962).
[16] P. A. Alekseev, V. N. Lazukov, R. Osborn *et al.*, RAL Annual Report, ISIS 95, A50 (1995).

[17] U. Walter, J. Phys. Chem. Solids **45**, 401 (1984).
[18] K. W. H. Stevens, Proc. Phys. Soc. London, Sect. A **65(3)**, 209 (1952).
[19] D. J. Newman and B. Ng, Rep. Prog. Phys. **62**, 699 (1989).
[20] E. S. Klement'ev, Preprint IAÉ 5822/9 (1994).
[21] P. A. Alekseev, E. S. Klement'ev, V. N. Lazukov *et al.*, JETP Lett. **63**, 1000 (1996).
[22] P. A. Alekseev, W. Bührer, V. N. Lazukov *et al.*, Physica B **217**, 241 (1996).
[23] P. A. Alekseev, V. N. Lazukov, A. Yu. Rumyantsev *et al.*, J. Magn. Magn. Mater. **75**, 323 (1988).
[24] P. A. Alekseev, A. Andreeff, H. Griessman *et al.*, Phys. Status Solidi B **27**, 87 (1980).
[25] E. A. Goremychkin, É. Myule, I. Natkanets *et al.*, Preprint OIYaN R14-84-632 (1984).
[26] A. P. Murani, R. Raphel, Z. A. Bowden *et al.*, Phys. Rev. B **53**, 8188 (1996).
[27] A. P. Murani, *Theoretical and Experimental Aspects of Valence Fluctuations and Heavy Fermions*, L. Gupta and S. K. Malic (eds.), Plenum Press, New York (1987).

Translated by Paul F. Schippnick

1/*f* Noise in a nonequilibrium phase transition: experiment and mathematical model

V. P. Koverda,*) V. N. Skokov, and V. P. Skripov

Institute of Thermal Physics, Russian Academy of Sciences, Ural Branch, 620219 Ekaterinburg, Russia
(Submitted 23 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1748–1757 (May 1998)

The results of an experimental investigation of a high-power source of broad-band $1/f$ noise, which can be generated in a system of two interacting nonequilibrium phase transitions, are presented. This process takes place when a normal conductor-superconductor phase transition is superposed on the critical liquid-vapor transition in a boiling coolant. A mathematical model describing a nonequilibrium phase transition in a complicated nonlinear system with two interacting order parameters, which involves the conversion of white noise into stochastic fluctuations of the order parameters with $1/f$ and $1/f^2$ spectra, is proposed. The properties of the model fluctuations with a $1/f$ spectrum agree qualitatively with the experimentally observed properties. A characteristic difference between the model fluctuations with a $1/f^2$ spectrum and random walks is also noted. © *1998 American Institute of Physics.* [S1063-7761(98)01305-5]

1. INTRODUCTION

The interest in random processes whose spectral density varies in inverse proportion to the frequency, which has not waned over the course of many years, is due to the widespread occurrence of this phenomenon and the absence of generally accepted, universal mathematical models. Stochastic processes with a spectrum that is inversely proportional to the frequency ($1/f$ noise) are observed in systems of very diverse types, i.e., from current fluctuations in radiophysical devices to cellular automata simulating the phenomenon of self-organized criticality, and nonlinear dynamic systems with intermittency. As a result of numerous investigations (see, for example, Refs. 1–5), the main properties of $1/f$ noise have been established. However, the mechanism for the appearance of a $1/f$ spectrum and the location of its sources are frequently unclear. Thermal mechanisms for the appearance of $1/f^n$ noise in metals were discussed in Refs. 6 and 7. It was noted that the value of n depends on the thermal isolation of the substrate. Theoretical arguments regarding the possibility of switching off the heat conduction mechanism were advanced in Ref. 8. It was claimed therein that $1/f^n$ noise arises as a result of a nonlinear interaction between diffusive and heat conduction modes. The idea that n can vary in the passage through a second-order phase transition point was advanced in Ref. 9. General arguments regarding the appearance of $1/f^n$ noise in distributed systems were presented in Ref. 10. According to Ref. 10, the scale-invariant power-law form of the spectra in phonon systems is attributable to fluctuations of the phase diffusion rate and relaxation of the phonon modes. Most of the published studies devoted to this phenomenon deal with spatially distributed systems. In Ref. 11 we presented the results of the experimental detection of the generation of signals with a $1/f$ spectrum when a superconductor undergoes Joule heating in a boiling coolant. The distinguishing feature of these experiments is that there was only one source of stochastic signals with a $1/f$ spectrum in the system. Such behavior is associated with the interaction of phase transitions occurring in nonlinear subsystems containing both a superconductor with a current and a boiling coolant.

In this paper we present detailed experimental results from the observation of the stochastic process with a high-intensity $1/f$ spectrum accompanying the Joule heating of thin films of high-temperature superconductors in boiling nitrogen and propose a mathematical model of a nonequilibrium phase transition in a two-dimensional system which converts white noise into noise with $1/f$ and $1/f^2$ spectra.

2. EXPERIMENT

The experiments were performed on thin-film $YBa_2Cu_3O_{7-x}$ bridges. The thickness of the bridges was ≈ 0.3 μm, its width was 0.5–1.0 mm, and its length was 2–8 mm. The superconducting transition temperatures were $T_c = 86–88$ K, and the critical current densities were $j_c = 10^5–10^6$ A/cm^2 at 77 K. During the experiment the samples were immersed directly into liquid nitrogen. The current–voltage characteristics were measured in a potentiostatic regime.

When a transport current above the critical value was passed, thermal instability developed in the films with the formation of a temperature–electric-field domain, leading to the appearance of ''descending'' segments on the current–voltage characteristic with a negative differential resistance. Figure 1 presents a typical current–voltage characteristic of a sample. Segment AB corresponds to a superconducting or stable low-resistivity state. Segment BC corresponds to the appearance of a thermal domain in the superconducting film. Heat exchange with the liquid nitrogen is mediated on this segment by single-phase convection. When there was free-convective heat transfer, the nitrogen layer near the surface of the high-T_c superconductor heating element was superheated. When superheating amounting to several degrees was achieved, local bubbling of nitrogen on a ''weak'' spot of the sample occurred. Bubbling was observed visually, and

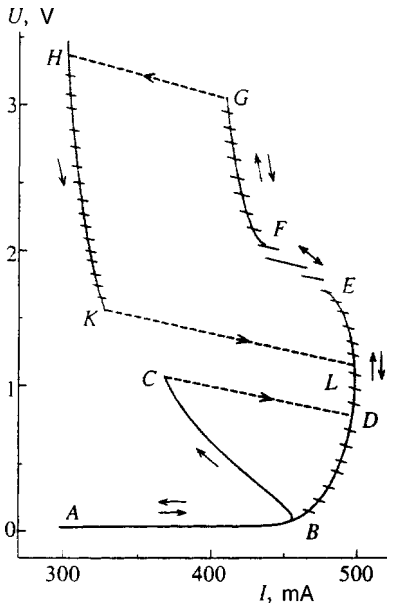

FIG. 1. Current–voltage characteristic of a thin-film high-T_c superconductor bridge in boiling nitrogen.

a jump occurred along the loading line on the current–voltage characteristic (*CD* in Fig. 1). Segment *DE* in Fig. 1 corresponds to local (''point'') boiling.

Oscillations associated with oscillations of the heat-removal rate were observed on the current–voltage characteristic in a nucleated boiling regime. In Fig. 1 they are indicated by short strokes (along the loading line of the electrical circuit). The fluctuations of the voltage drop on the potential contacts and the transport current were recorded by S9-8 oscilloscopes connected to a computer. At certain loads (segment *EF* in Fig. 1) a significant increase in the oscillation amplitude was observed, and random jumps occurred from point *E* to point *F* (along the loading line). The detachment diameter of the bubbles increased, and the train of individual bubbles transformed into a vapor jet with the formation of a ''dry'' spot at the boiling site. The heat flux density in this case corresponded to the first critical flux density for the transition of nitrogen from nucleated to film boiling ($\sim 10^6$ W/m^2). In other words, segment *EF* in Fig. 1 can be associated with a nonequilibrium phase transition from nucleated to film boiling (a boiling crisis). However, the local nature of the boiling process was not destroyed, and the vapor was generated, as before, from a focal ''weak'' spot on the film. The local nature of the boiling crisis observed can be associated with the strong variation of the release of heat along the sample due to localization of the thermal domain in a potentiostatic regime.

When the load was increased further, the vapor boundary underwent abrupt motion with the formation of a continuous vapor film on the sample surface (point *G* in Fig. 1). The reverse transition from continuous film boiling to nucleated boiling occurred at different loads (point *K*), i.e., hysteresis was observed on the current–voltage characteristic.

The measured oscillograms of the voltage drop on the potential contacts (or the transport current) were used to calculate the spectral densities of the oscillations associated

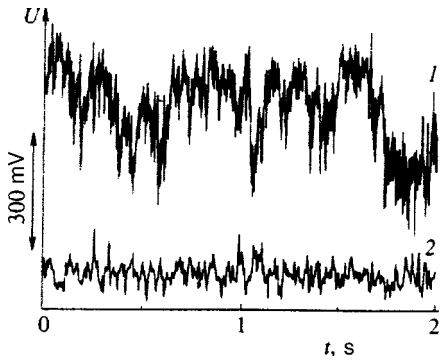

FIG. 2. Realization of the voltage drop on potential contacts in the critical regime (*1*) and in the ''ordinary'' nucleated boiling regime (*2*).

with the boiling of nitrogen. With the exception of segment *EF* in Fig. 1, the spectra obtained had a typical Lorentzian shape. On segment *EF*, along with the nonequilibrium superconductor–normal-conductor phase transition, there was a phase transition associated with a local boiling crisis under the conditions of Joule heating (which always occurs when a current–voltage characteristic has ''descending'' segments). Figure 2 (curve *1*) shows a typical realization of the voltage drop on the potential contacts corresponding to segment *EF* in Fig. 1 (the critical regime). For comparison, the figure also shows a realization of the voltage drop (curve *2*) corresponding to segment *DE* in Fig. 1. The length of each realization was determined by the buffer memory of the oscilloscope and corresponded to 2048 points. The time interval between the points varied from 500 μs to 20 ms, making it possible to vary the signal observation time (which determines the lower frequency boundary in the spectrum) from approximately 1 to 40 s. The observed oscillatory processes were stationary. Figure 3 presents typical distribution functions of the voltage-drop fluctuation amplitude in the critical regime. It is seen from the figure that the process observed can be considered Gaussian in some approximation. However, it is seen that the experimental distribution function could be separated into two maxima. The separation is greater, the greater is the angle of inclination of the loading line of the electrical circuit relative to the current–

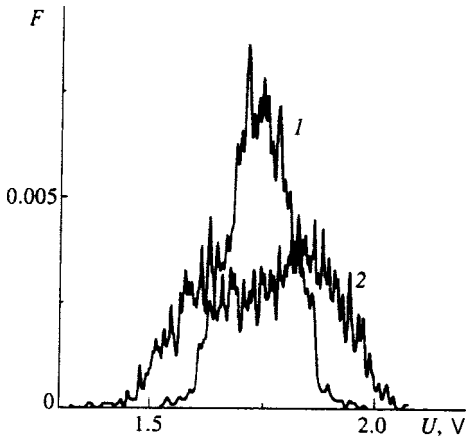

FIG. 3. Distribution functions of the voltage-drop fluctuation amplitude: *1*—load resistance $R_0 = 1\ \Omega$, *2*—$R_0 = 5\ \Omega$.

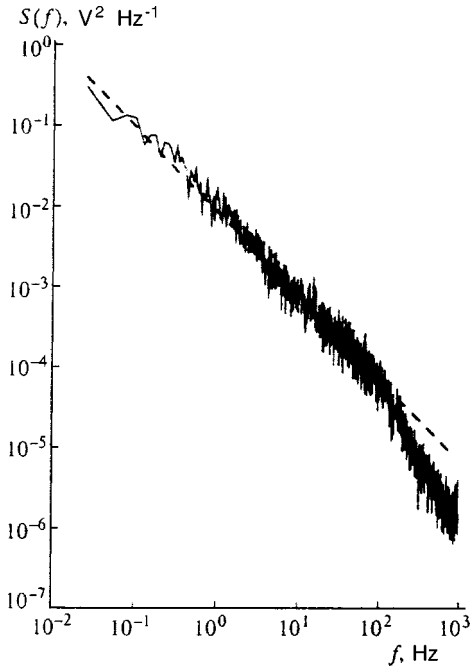

FIG. 4. Spectrum determined from experimental realizations in the critical regime.

voltage characteristic (the greater is the load resistance of the electrical circuit). Complete symmetry of the distribution function relative to its mean value was observed only when the working point corresponded to the midpoint of the interval EF (see Fig. 1). Otherwise the distribution function was asymmetric, i.e., the system spent more time at point E or at point F.

The spectral densities were calculated from the measured realizations by fast Fourier transformation. For all the realizations recorded in the critical regime with different time intervals between the points, the spectral density varied in inverse proportion to the frequency in the frequency range $f \leqslant 100$ Hz. Such behavior remained unchanged when the voltage of the source varied over a range of several tenths of a volt and when the load resistance of the circuit varied by several fold, which corresponded to variation of the angle of inclination of the loading line relative to the current–voltage characteristic over a range of $\simeq 30°$. It should be noted that the observed amplitudes of the voltage drop on the sample in the critical regime amounted to tenths of a volt and were thus more than five orders of magnitude greater than the characteristic signal amplitudes detected when such $1/f$ noise is observed in solids.

Figure 4 presents the spectral density of the voltage-drop fluctuations in the critical regime. The spectrum was obtained from realizations recorded with different temporal discretization and was averaged over a set of realizations (about 70 realizations). The dashed line in Fig. 4 was obtained as a result of treatment of the low-frequency part of the spectrum ($f = 2\times 10^{-2} - 10^2$ Hz) and corresponds to a function $\propto 1/f^{1.02}$. In the frequency range $f > 10^2$ Hz a dependence close to $1/f^{1.5}$ was observed. We stress that Fig. 4 corresponds to a fairly large ensemble of realizations; however, the spectra for the individual realizations (which cover three orders of magnitude in the frequency) also had the $1/f$ form.

Thus, the interaction of two nonequilibrium phase transitions associated with nonlinear processes in a superconductor with a current and a boiling coolant gives rise to intense stochastic fluctuations with a $1/f$ spectrum.

3. MATHEMATICAL MODEL

The experiment described certainly shows that we are dealing here with a single point source of $1/f$ stochastic signals. These results cannot be explained within the widely used model of an exponentially broad distribution of relaxation times,[1–3] in which $1/f$ noise is treated as a result of the superposition of random relaxation processes. To obtain a range covering four orders of magnitude with respect to the frequency, a spread of time constants of the individual sources over a range of six orders of magnitude is required according to that model.[3]

An alternative approach to explaining the experimental results described is to find a dynamic system which would convert white noise into noise with a $1/f$ spectrum. The experiment also showed that such a system operates in the vicinity of a nonequilibrium phase transition.

In the vicinity of the phase transition, for the order parameter T we can write the equation

$$\dot{T} = \alpha T - \beta T^3 + \Gamma(t), \tag{1}$$

where $\Gamma(t)$ is a source of white noise, which is present in any physical system. It is known that the fluctuations described by an equation of the type $\dot{x} = F(x) + \Gamma(t)$, where $F(x)$ is a smooth continuous function and $\Gamma(t)$ is Gaussian δ-correlated noise, cannot lead to a $1/f$ spectrum. Equations of this type give a spectrum with a Lorentzian or quasi-Lorentzian shape. In particular, Eq. (1) gives a Lorentzian spectrum.

As was noted above, $1/f$ behavior was observed experimentally under conditions in which two interacting nonequilibrium phase transitions take place in a two-component nonlinear system. Therefore, the following two-dimensional system of equations was proposed in Ref. 12 for this case:

$$\dot{T} = -\beta T P^2 + \alpha P + \Gamma(t),$$

$$\dot{P} = -\beta P T^2 + \gamma T + \Gamma(t). \tag{2}$$

System (2) describes two phase transitions with the interacting order parameters T and P. The parameters T and P are determined by the specific nature of the nonequilibrium phase transitions and can be related to the temperature, the dimensions of the dissipative regions, or fluctuations of the kinetic coefficients. Fluctuations of the latter near phase-transition points were treated theoretically in Ref. 9. We shall describe the properties of this system without specifying the physical meaning of T and P in the present work.

When the coefficients are equal ($\alpha = \gamma$), system (2) has a potential character and is equivalent to Eq. (1). In this case it gives coinciding Lorentzian spectra for T and P regardless of the value of the coefficients. However, if $\alpha < \gamma$ holds, a noise intensity level $\Gamma(t)$ at which the fluctuations of T are proportional to $1/f$ and the fluctuations of P are proportional

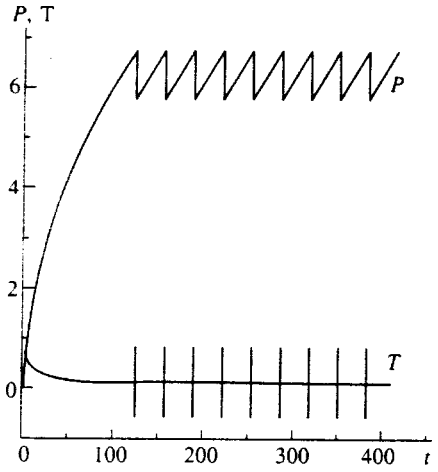

FIG. 5. Results of the numerical integration of system (4).

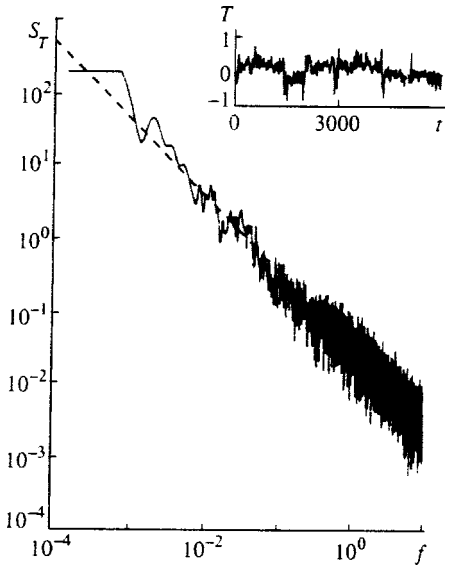

FIG. 6. Spectrum of $S_T(f)$ and realization of $T(t)$ obtained by numerical integration of the system (6). Dashed line—plot of $S_T(f) \propto 1/f^{1.05}$.

to $1/f^2$ can easily be selected for any values of the coefficients. Such a dependence was traced in Ref. 12 over four to five orders of magnitude in frequency by performing numerical integration by the Euler method and employing a sequence of random numbers with a Gaussian distribution as $\Gamma(t)$.

Let us first consider the behavior of the system (2) in the absence of noise $[\Gamma(t)=0]$. The asymptotic solutions in the limit $t\to\infty$ for $\beta>0$, $0<\alpha<\gamma$, and nonzero initial conditions have the form

$$T=\sqrt{\frac{\alpha}{2\beta(\gamma-\alpha)t}}, \quad P=\sqrt{\frac{2\alpha(\gamma-\alpha)t}{\beta}}, \tag{3}$$

i.e., they describe a relaxation process with $T\propto t^{-1/2}$ and $P\propto t^{1/2}$, so that their product is a constant: $TP=\alpha/\beta$.

However, numerical integration of the system by the Euler method with the integration step dt,

$$\begin{aligned} T_{i+1}&=T_i+(-\beta T_i P_i^2+\alpha P_i)dt,\\ P_{i+1}&=P_i+(-\beta P_i T_i^2+\gamma T_i)dt, \end{aligned} \tag{4}$$

reveals that as time passes T and P follow the asymptotes (3) only up to a certain limit, which depends only on dt. When this limit is achieved, P decreases abruptly, while T increases, then relaxation occurs again, and thereafter the process repeats itself (Fig. 5). Such behavior is attributed to the fact that the derivative is taken at the initial point of the ith interval when the system is integrated numerically and the numerical integration is unstable. To make it more stable we went over to the system

$$\begin{aligned} T_{i+1}&=T_i+\{[-\beta((1-c)T_i+cT_{i+1})]P_i^2+\alpha P_i\}dt,\\ P_{i+1}&=P_i+\{[-\beta((1-c)P_i+cP_{i+1})]T_i^2+\gamma T_i\}dt, \end{aligned} \tag{5}$$

in which, depending on c $(0\leqslant c\leqslant 1)$, the value of T_i in the first equation and the value of P_i in the second equation can be taken for calculating the derivative not only at the initial $(c=0)$ or final point $(c=1)$, but also at the midpoint of the ith integration interval. In particular, when $c=1$, the system (5) gives the asymptotes (3) for any value of dt and initial conditions.

However, if the value of the quantity being integrated is taken away from the initial point of the integration interval, some ambiguity in the way to introduce the noise into the system appears. We tested several variants, and below in this paper we shall present the results of the numerical integration of a system with noise of the following form:

$$\begin{aligned} T_{i+1}&=\{T_i+[-(1-c)\beta T_i P_i^2+\alpha P_i]dt\}\\ &\quad\times(1+cP_i^2dt)^{-1}+W_i dt,\\ P_{i+1}&=\{P_i+[-(1-c)\beta P_i T_i^2+\gamma T_i]dt\}\\ &\quad\times(1+cT_i^2dt)^{-1}+W_i dt, \end{aligned} \tag{6}$$

in which the value of the derivative is calculated during the numerical integration precisely as in the system without noise (5) and the effect of the noise is taken into account later on. The noise $W_i dt$ has the form of a random-walk differential, where W_i is a Gaussian distribution of random numbers with the root-mean-square deviation D.

As an example, Figs. 6–8 present the results of the numerical integration of the system (6) with the following values of the parameters: $\beta=1$, $\alpha=0.5$, $\gamma=1$, $D=2.14$, and $c=0.5$. When $dt=0.05$, each integration interval includes 130 000 points. A realization of the process $T(t)$ and its spectrum $S_T(f)$ are presented in Fig. 6, and the distribution function F_T is presented in Fig. 7. It is seen that it resembles a Gaussian distribution, but this is true only when $\alpha<0.5\gamma$. When $\alpha>0.5\gamma$ holds, F_T has a minimum near zero, which is characteristic of other values of the parameters in the realizations of (6) calculated.

For comparison, the inset in Fig. 8 presents a realization of the random-walk process

$$w(t)=\int_0^t W(t)dt,$$

which has the same spectrum as $S_P(t)$, together with a realization of $P(t)$. This comparison reveals that the $1/f^2$ diver-

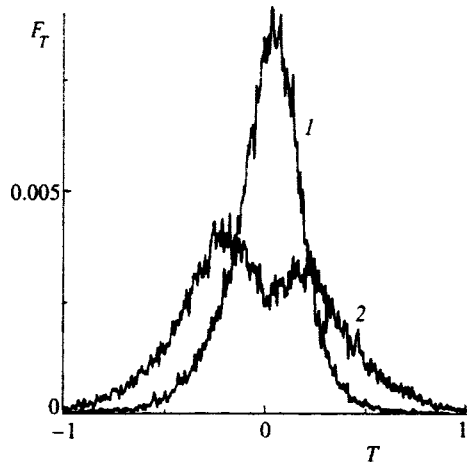

FIG. 7. Distribution function of the fluctuation amplitude of $T(t)$ obtained by numerical integration of the system (6): *1*—$\alpha=0.5$, *2*—$\alpha=0.75$.

gence of the spectrum as $f\to 0$ for the realization of $P(t)$ is caused by increases in the derivatives, rather than an increase in the amplitude as in the case of random walks. The same applies to $T(t)$, which has $1/f$ divergence. Both noise components, i.e., the $1/f$ and $1/f^2$ components, are stationary and do not depend on the initial conditions. Good reproducibility of the spectra and the distribution functions is noted for other sequences of random numbers w_i. A decrease in c from unity down to zero does not strongly alter the appearance of the realizations, but a peak appears in the high-frequency region of their spectra due to the instability of the numerical integration at small values of c.

It is noteworthy that the spread of the fluctuation amplitudes of P and T for an assigned value of dt correlates with the limiting values of P and T, which are given by the integration of the system (4) without noise. The spikes of largest absolute value only slightly exceed these limits. Another interesting property of the solutions of the system is the fact that the product $P(t)T(t)$ has a spectrum of white noise, and the mean value $\langle P(t)T(t)\rangle \simeq \alpha/\beta$ holds to within good accuracy for any values of the parameters of the system and realization lengths. This means that the equation for the fluctuations of T adjusts itself to the characteristic zero eigenfrequency and that we are dealing with self-organized criticality.

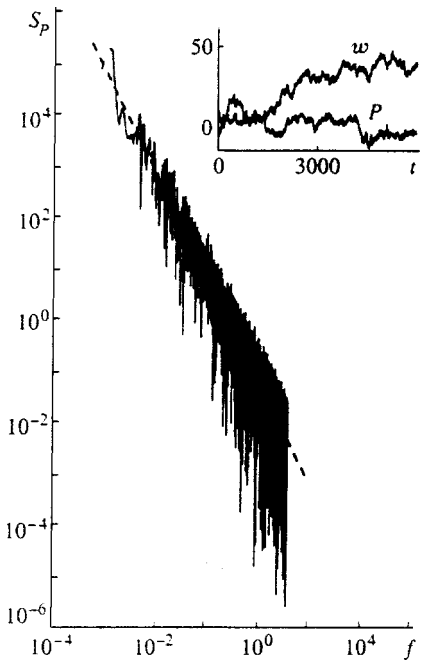

FIG. 8. Spectrum of $S_P(f)$ and realizations of $P(t)$ and of the random walks $w(t)$ obtained by numerical integration. Dashed line—plot of $S_P(f)\propto 1/f^2$.

The results of the integration of (6) for the parameters indicated with $dt=0.05$ and the root-mean-square deviation $D=1.8$ in Figs. 6–8 show the behavior of the $1/f$ noise and $1/f^2$ noise over a range of five orders of magnitude in frequency. When the number of integration points is increased further, the branches of the fluctuation spectra of T and P become horizontal, i.e., for assigned values of the integration step and the dispersion of the $1/f$ and $1/f^2$ spectra are observed only in a limited frequency range. However, this range can be increased, and more distant low-frequency $1/f$ and $1/f^2$ asymptotes can be traced for the same values of the parameters α, β, and γ of the system, if the integration step dt is diminished by a factor of n with a simultaneous increase in D by a factor of $\sqrt{n}$ as the number of intervals is increased. Therefore, the hope remains that if we would have a source of ''true'' white noise in the form of a sequence of δ functions with infinitely short spacing and if we knew how to integrate such equations, the system (2) would give branches of the $1/f$ and $1/f^2$ spectra with infinite low-frequency asymptotes down to $f=0$.

At the values $0<\alpha<\gamma$ the proposed model gives a $1/f$ spectrum for the fluctuations of T, a realization of $T(t)$, and a distribution function which conform with the experiment described above (Figs. 2–4). The properties of the $1/f$ noise in this case are similar to those observed in radiophysical devices, where the mean is $\langle T(t)\rangle=0$. However, if α and γ are negative $(0>\alpha>\gamma)$, or if they are positive, but the noise acts on the equations for P and T in antiphase, we have $\langle T(t)\rangle\neq 0$ and $\langle P\rangle\neq 0$. The behavior of the $1/f$ and $1/f^2$ components and their stationary character are maintained in this case for the same absolute values of α, β, and γ in the system. The realizations of the fluctuations of T in this case contain sharp spikes and outwardly resemble the realizations which are obtained in models of ''sand-pile'' cellular automata[13] or highway traffic,[14] in which there has recently been interest in connection with the study of self-organized criticality.

4. CONCLUSIONS

In this paper we have presented the results of an experimental investigation of an unconventional source of $1/f$ noise for radiophysics, where we are apparently always dealing with the superposition of a large number of sources of noise of different nature. In our case the source of intense broad-band $1/f$ noise is associated with the interaction of two nonequilibrium phase periods. The mathematical model proposed for describing its properties conforms qualitatively with the experiment, but, apart from the generation of $1/f$

noise, it predicts the generation of a satellite with a $1/f^2$ spectrum. The properties of the $1/f^2$ noise differ from those of random walks, since no increase in the amplitude of the spikes is observed, and the divergence, as in the case of the $1/f$ noise, is associated with increases in the derivatives. The $1/f^2$ noise was not observed experimentally. It is possible that the fluctuations responsible for that noise are associated with the parameters of the nonequilibrium phase transition in the boiling coolant, require additional diagnostics, and were, therefore, not detected. In any case, this poses a new problem for experimentation.

This work was supported by the Russian Fund for Fundamental Research (Project Code 96-02-16077a).

*)E-mail: iva@itp.e-burg.su

[1] Sh. M. Kogan, Usp. Fiz. Nauk, **145**, 285 (1985) [Sov. Phys. Usp. **28**, 170 (1985)].
[2] G. P. Zhiga'lskiĭ, Usp. Fiz. Nauk **167**, 623 (1997) [Phys. Usp. **40**, 599 (1997)].
[3] M. J. Buckingham, *Noise in Electronic Devices and Systems*, E. Horwood, Chichester; Halsted Press, New York (1983) [Russ. transl., Mir, Moscow (1986)].
[4] Yu. L. Klimontovich, *Statistical Theory of Open Systems, Vol. 1: A Unified Approach to Kinetic Description of Processes in Active Systems*, Kluwer Academic Publishers, Dordrecht–Boston (1995).
[5] Yu. S. Levitan, N. N. Panchenko, and O. A. Sinkevich, Dokl. Akad. Nauk SSSR **302**, 1359 (1988) [Sov. Phys. Dokl. **33**, 750 (1988)].
[6] P. Dutta, P. Dimon and P. M. Horn, Phys. Rev. Lett. **43**, 646 (1979).
[7] P. Dutta and P. M. Horn, Rev. Mod. Phys. **53**, 497 (1981).
[8] R. O. Zaĭtsev, JETP Lett. **58**, 915 (1993).
[9] R. O. Zaĭtsev, Zh. Éksp. Teor. Fiz. **90**, 1288 (1986) [Sov. Phys. JETP **63**, 752 (1986)].
[10] Yu. E. Kuzovlev, Zh. Éksp. Teor. Fiz. **111**, 2086 (1997) [JETP **84**, 1138 (1997)].
[11] V. P. Koverda, V. N. Skokov, and V. P. Skripov, JETP Lett. **63**, 775 (1996).
[12] V. P. Koverda, V. N. Skokov, and V. P. Skripov, Dokl. Akad. Nauk SSSR **356**, 614 (1997) [Phys. Dokl. **42**, 532 (1997)].
[13] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. A **38**, 364 (1988).
[14] Kh. Zhang and G. Hu, Phys. Rev. E **52**, 4664 (1995).

Translated by P. Shelnitz

Dynamics of small spin polaron in the three-band model of two-dimensional spherically symmetric antiferromagnet

A. F. Barabanov and E. Žąsinas

Institute for High Pressure Physics, 142092 Trojtsk, Moscow Region, Russia

R. O. Kuzian

Institute for Materials Science, 252180 Kiev, Ukraine

L. A. Maksimov

Russian Research Center, Kurchatov Institute, 123182 Moscow, Russia

(Submitted 26 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 1758–1777 (May 1998)

The retarded Green's function $G(\mathbf{k},\omega)$ of a single small spin polaron in the three-band model for the CuO_2 plane is calculated in the self-consistent Born approximation. It is shown that such a spin polaron is a good quasiparticle excitation for realistic values of spin exchange J and effective hopping τ. The polaron spectral density $A_p(\mathbf{k},\omega)$ demonstrates small damping in contrast to the results of calculations starting from the bare hole; i.e., the pole strength $Z_p(\mathbf{k})$ of the energetically low-lying quasiparticle peak varies from 50% to 82% for $J/\tau \sim 0.1-0.7$. The quasiparticle peak dispersion reproduces the main features of the bare polaron spectrum $\Omega_{\mathbf{k}}$ near the band bottom. The spherically symmetric approach is used for the description of spin excitations. This approach makes it possible to consider the quantum antiferromagnetic background without the spontaneous symmetry breaking and the unit cell doubling. The new method of the self-consistent calculation, which is based on continuous-fraction expansion of the Green's function, is presented in detail. The method preserves the proper analytical properties of the Green's function and makes it possible to analyze the nature of its singularities.
 [S1063-7761(98)01405-X]

1. INTRODUCTION

The hole motion in a two-dimensional (2D) $s=1/2$ quantum antiferromagnet (AFM) has been studied in depth theoretically.[1] The important question is whether a hole injected in the undoped ground state behaves like a quasiparticle. This problem is mainly investigated in the framework of self-consistent Born approximation (SCBA) for the $t-J$ model[2–7] and Kondo lattice.[8] There are only a few studies devoted to the three-band Hubbard model or the Emery model,[9,10] which is more realistic for CuO_2 planes in high-T_c superconductors (HTSC). For the $t-J$ model it was shown that the spectral density function $A_h(\mathbf{k},\omega)$ of a doped hole revealed a quasiparticle peak of intensity $Z_{\mathbf{k}} \approx J/t$ and a broad incoherent part that has a width of about $(6-7)t$. The quasiparticle band bottom corresponds to the momenta $\mathbf{k}_1=(\pm\pi/2,\pm\pi/2)$. Similar results were obtained for the Emery model.[11,12] The presence of a large incoherent part and small intensity of the quasiparticle peak indicate that bare holes are rather poor elementary excitations even for the case in which $\mathbf{k}$ is close to $\mathbf{k}_1$.

In order to investigate the hole motion in the $t-J$ model one usually decouples the hole operator into a spinless fermion and an antiferromagnetic magnon operator. As a result, the zero approximation corresponds to the dispersionless band with zero energy of the hole. The hopping of the particle appears only due to the fermion–magnon scattering, which is treated by the usual perturbation method in $\mathbf{k}$-space. For this reason, we think that in this approach the resulting quasiparticle pole in the fermion Green's function involves mainly a polaron with a large radius. A similar situation takes place in the usual treatment of a hole motion in the effective three-band model[11–13] and the Kondo-lattice model,[8] where one starts from a bare hole rather than from a small-radius magnetic polaron.

In the framework of the effective three-band model we studied the spectral density function $A_p(\mathbf{k},\omega)$ of a single small polaron, i.e., an excitation which at the outset takes into account a local hole-spin coupling. It is known that the simplest candidate for such a small polaron is an analog of the so-called Zhang–Rice singlet in the CuO_4 plaquette.[14,15] The mean-field spectrum $\Omega_{\mathbf{k}}$ of this excitation has been studied extensively[15] and will be used as the zero approximation in our treatment. We shall consider the coupling of a small polaron to spin-wave excitations in SCBA for the corresponding two-time retarded Green's function $G(\mathbf{k},\omega)$.

Our motivations to study $A_p(\mathbf{k},\omega)$ and the corresponding quasiparticle band are the following. First, it is easy to show for the one-hole problem that the mean-field energy of the polaron $\Omega_{\mathbf{k}}$ represents the center of gravity of the spectral function:

$$\Omega_k = \int_{-\infty}^{\infty} \omega A_p(\mathbf{k},\omega)\, d\omega. \tag{1}$$

This means that the minimum of $\Omega_{\mathbf{k}}$ is the upper bound of the actual position for the quasiparticle band bottom. The SCBA based on a bare hole Green's function gives the minimum value of the quasiparticle energy $\omega_h^{\min}=-2.6\tau$[12] for a typical value of copper–copper AFM exchange constant $J=0.7\tau$. Here τ is a constant of the effective oxygen–oxygen hopping via an intervening copper site (note that our unit of energy is twice that of Ref. 12, $\tau=2t$). As to the value of the small polaron mean-field band bottom, it turns out to be substantially lower than $\omega_h^{\min}$, $\Omega_{\mathbf{k}}=-3.17\tau$, for the same value of J/τ. We may conclude, therefore, that important local correlations are lost in SCBA when we start from the bare hole operators.

Second, we shall show that a small polaron represents the elementary hole excitation much better than a bare hole dressed by magnons in the framework of the SCBA. This is manifested by a relatively large intensity of a quasiparticle peak in our calculation.

Finally, the mean-field spectrum $\Omega_{\mathbf{k}}$ of the simplest small spin polaron explicitly depends on the state of the antiferromagnetic background. In the case of the long-range-order state $\Omega_{\mathbf{k}}$ demonstrates a flat band region, which is close to the magnetic Brillouin zone boundary.[15] This region corresponds to the bottom of the band. Moreover, if the direct oxygen–oxygen hopping, the finite temperature, and a more complicated form of a small polaron wave function are taken into account, then $\Omega_{\mathbf{k}}$ reproduces the experimentally observed, extended saddle point,[16–20] which is directed along the line $(0,\pi)-(0,0)$.[21] Therefore, it seems important to ascertain whether the quasiparticle band reproduces the peculiarities of $\Omega_{\mathbf{k}}$ dispersion. Using a very simple variant of the model, we shall determine below whether this is in fact the case.

The distinctive feature of our investigation is the consideration of the AFM copper spin subsystem in a spherically symmetric approach.[22,23] Such an approach is most appropriate in treating the quantum 2D AFM at any finite temperature. As a result, the scattering of a spin polaron by spin excitations in the singlet spin background leads to the spectral function periodicity relative to the full Brillouin zone. Note that the conventional two-sublattice spin approach leads to periodicity relative to the magnetic (reduced) Brillouin zone.[2–7,12]

This paper is organized as follows. In Sec. 2 we give the derivations for the self-consistent equation for the Green's function in the case of the small polaron approach. In Sec. 3 we present the procedure that makes it possible to avoid the iterative solution of the self-consistent equation for complex energies. The procedure is based on the continuous-fraction expansion of the Green's function and consequently makes it possible to calculate the coefficients of the continuous-fraction expansion with use of the quadrature method. In Sec. 4 we deal with the termination of the continuous-fraction, which leads to the correct analytical properties of the resulting Green's function. The numerical results for the self-energies and spectral functions, the relation of our results to the previous approaches, and the discussion are given in Sec. 5. In Sec. 6 we summarize the results. The Appendix contains some details of the approach, which gives the expression for the integrals over the spectral density in terms of the chain representation of the continuous fraction.

Some of our results were presented in a brief report.[24] In this paper we present additional results, describe the new method, and give more details about the calculations.

2. EFFECTIVE HAMILTONIAN AND SMALL POLARON GREEN'S FUNCTION

Following Refs. 9, 10, and 15, we adopt the Hamiltonian that corresponds to one-hole problem in the CuO_2 plane of the high-T_c superconductors:

$$\hat{H}=\tau\sum_{\mathbf{r},\mathbf{a}_1,\mathbf{a}_2,\sigma,\sigma'} c^{\dagger}_{\mathbf{r}+\mathbf{a}_1,\sigma}c_{\mathbf{r}+\mathbf{a}_2,\sigma'}\left(\frac{1}{2}\,\delta_{\sigma\sigma'}+2\mathbf{s}_{\sigma\sigma'}\mathbf{S}_{\mathbf{r}}\right)+\frac{J}{2}\sum_{\mathbf{r},\mathbf{g}}\mathbf{S}_{\mathbf{r}}\mathbf{S}_{\mathbf{r}+\mathbf{g}}, \tag{2}$$

where $\mathbf{a}_1,\mathbf{a}_2=\pm\mathbf{g}_x/2,\ \pm\mathbf{g}_y/2$, $\mathbf{g}=\pm\mathbf{g}_x,\ \pm\mathbf{g}_y$. Here and below $\mathbf{g}_{x,y}$ are the basic vectors of a copper square lattice ($|\mathbf{g}|\equiv 1$), $\mathbf{r}+\mathbf{a}$ are four vectors of the O sites nearest to the Cu site $\mathbf{r}$, the operator $c^{\dagger}_{\sigma}$ creates a hole with the spin index $\sigma=\pm 1$ at the O site, $\mathbf{s}_{\sigma\sigma'}=\boldsymbol{\sigma}_{\sigma\sigma'}/2$, and the operator $\mathbf{S}$ represents the localized spin at the copper site. As mentioned above, τ is the integral of oxygen hole hoppings, which takes into account the coupling of the hole motion with copper spin subsystem, and J is the constant of the nearest-neighbor AFM exchange between the copper spins.

It is well known that the most prominent feature of the Hamiltonian (2) is that the low-energy physics of hole excitations is described by the Bloch sums $\mathscr{B}^{\dagger}_{\mathbf{k},\sigma}$ which are based on one-site small-polaron operators $\mathscr{B}^{\dagger}_{\mathbf{r},\sigma}$

$$\mathscr{B}^{\dagger}_{\mathbf{k},\sigma}=\frac{1}{\sqrt{NK_{\mathbf{k}}}}\sum_{\mathbf{r}} e^{i\mathbf{k}\mathbf{r}}\mathscr{B}^{\dagger}_{\mathbf{r},\sigma}, \tag{3}$$

$$\mathscr{B}^{\dagger}_{\mathbf{r},\sigma}=\frac{1}{2}\sum_{\mathbf{a}}\left(c^{\dagger}_{\mathbf{r}+\mathbf{a},\sigma}Z^{\overline{\sigma}\sigma}_{\mathbf{r}}-c^{\dagger}_{\mathbf{r}+\mathbf{a},\overline{\sigma}}Z^{\sigma\overline{\sigma}}_{\mathbf{r}}\right). \tag{4}$$

$$K_{\mathbf{k}}=\left\langle\frac{1}{N}\sum_{\mathbf{r},\mathbf{r}'} e^{-i\mathbf{k}(\mathbf{r}-\mathbf{r}')}\{\mathscr{B}_{\mathbf{r},\sigma},\mathscr{B}^{\dagger}_{\mathbf{r}',\sigma}\}\right\rangle=1+\left(C_{\mathbf{g}}+\frac{1}{4}\right)\gamma_{\mathbf{k}}.$$

Here and below $\{,\}$ and $[,]$ stand for an anticommutator and commutator, respectively; $\langle\ldots\rangle\equiv Q^{-1}\mathrm{Tr}[e^{-\beta H}\ldots]$, and $Q=\mathrm{Tr}e^{-\beta H}$; $\beta=(kT)^{-1}$ is the inverse temperature; $\overline{\sigma}\equiv-\sigma$; $Z^{\sigma_1\sigma_2}_{\mathbf{r}}\equiv|\sigma_1\rangle\langle\sigma_2|$ are the Hubbard projection operators for Cu sites state, $\gamma_{\mathbf{k}}=(1/4)\Sigma_{\mathbf{g}}\exp(i\mathbf{k}\mathbf{g})$, and $C_{\mathbf{g}}=\langle\mathbf{S}_0\mathbf{S}_{\mathbf{g}}\rangle$.

To calculate the average for commutators and anticommutators such as $K_{\mathbf{k}}$, we take into account that these expressions are reduced to the two-site or three-site spin correlation functions. In principle, it is necessary to solve a self-consistent problem for the hole and spin subsystems in order to find these correlation functions. However, in the limit of a small number of holes it is possible to ignore the reverse influence of the holes on the spin subsystem. We can then use the results of Refs. 22 and 23, where the indicated spin

correlation functions are calculated in the spherically symmetric approach for the spin subsystem. In particular, in this approach the three-site correlation functions can be expressed in terms of the two-site spin correlation functions, $C_{\mathbf{r}}=\langle \mathbf{S}_0\mathbf{S}_{\mathbf{r}}\rangle$. We recall that due to the spherical symmetry $\langle S_i^{\alpha}S_j^{\beta}\rangle=\delta_{\alpha\beta}\langle S_i^{\alpha}S_j^{\alpha}\rangle=\frac{1}{3}\langle \mathbf{S}_i\mathbf{S}_j\rangle$, $\langle S_i^{\alpha}\rangle=0$. Simultaneously, as $T\to 0$, the spin subsystem is described by a long-range-order state with finite effective magnetization m, $C_{\mathbf{r}}(|\mathbf{r}|\to\infty)=m^2(-1)^{r_x+r_y}$; here the value of m is dictated by the Bose condensation of spin excitations at the antiferromagnetic vector $\mathbf{q}_0=(\pi,\pi)$.

Note that $\mathscr{B}^{\dagger}_{\mathbf{k},\sigma}|L\rangle$ corresponds to the CuO_2 plane state with the total spin equal to 1/2 if $|L\rangle$ is the singlet state. We treat $\mathscr{B}^{\dagger}_{k,\sigma}$ as a candidate for the elementary excitations operator and calculate the corresponding retarded two-time Green's function $G(\mathbf{k},\omega)$ and the spectral density

$$A_p(\mathbf{k},\omega)=-\frac{1}{\pi}\,\mathrm{Im}\,G(\mathbf{k},\omega+i0^+),$$

$$G(\mathbf{k},\omega)=\langle\langle \mathscr{B}_{\mathbf{k},\sigma}|\mathscr{B}^{\dagger}_{\mathbf{k},\sigma}\rangle\rangle_{\omega}\equiv -i\int_{t'}^{\infty}dt e^{i\omega(t-t')}\times\langle\{\mathscr{B}_{\mathbf{k},\sigma}(t),\mathscr{B}^{\dagger}_{\mathbf{k},\sigma}(t')\}\rangle. \tag{5}$$

Using the equations of motion, the retarded two-time Green's function $G(\mathbf{k},\omega)$ can be expressed (see, for example, Refs. 25 and 26) in the following form, which is analogous to the Dyson equation:

$$G^{-1}(\mathbf{k},\omega)=G_0^{-1}-\Sigma(\mathbf{k},\omega), \tag{6}$$

$$\Sigma(\mathbf{k},\omega)=\langle\langle\mathscr{R}|\mathscr{R}\rangle\rangle^{(irr)}=\langle\langle\mathscr{R}_{\mathbf{k},\sigma}|\mathscr{R}^{\dagger}_{\mathbf{k},\sigma}\rangle\rangle-\langle\langle\mathscr{R}_{\mathbf{k},\sigma}|\mathscr{B}^{\dagger}_{\mathbf{k},\sigma}\rangle\rangle\times\langle\langle\mathscr{B}_{\mathbf{k},\sigma}|\mathscr{B}^{\dagger}_{\mathbf{k},\sigma}\rangle\rangle^{-1}\langle\langle\mathscr{B}_{\mathbf{k},\sigma}|\mathscr{R}^{\dagger}_{\mathbf{k},\sigma}\rangle\rangle, \tag{7}$$

where

$$G_0=(\omega-\Omega_{\mathbf{k}})^{-1},\quad \mathscr{R}_{\mathbf{k},\sigma}=[\mathscr{B}_{\mathbf{k},\sigma},\hat{H}]=\frac{1}{\sqrt{NK_{\mathbf{k}}}}\sum_{\mathbf{r}}e^{-i\mathbf{kr}}\mathscr{R}_{\mathbf{r},\sigma}, \tag{8}$$

$$\mathscr{R}_{\mathbf{r},\sigma}=-4\tau\mathscr{B}_{\mathbf{r},\sigma}+\mathscr{R}^{\tau}_{\mathbf{r},\sigma}+\mathscr{R}^{J}_{\mathbf{r},\sigma}, \tag{9}$$

$$\mathscr{R}^{\tau}_{\mathbf{r},\sigma}=-\frac{\tau}{2}\,\sigma\Bigg(\sum_{\mathbf{g},\mathbf{a},\sigma_1}\sigma_1 Z_{\mathbf{r}}^{\bar{\sigma}\sigma_1}c_{\mathbf{r}+\mathbf{g}+\mathbf{a},\bar{\sigma}_1}-\sum_{\mathbf{g},\mathbf{a},\sigma_1,\sigma_2}\sigma_2 Z_{\mathbf{r}}^{\bar{\sigma}\sigma_1}Z_{\mathbf{r}+\mathbf{g}}^{\sigma_1\sigma_2}c_{\mathbf{r}+\mathbf{g}+\mathbf{a},\bar{\sigma}_2}\Bigg),$$

$$\mathscr{R}^{J}_{\mathbf{r},\sigma}=\frac{J}{4}\,\sigma\Bigg(\sum_{\mathbf{g},\mathbf{a},\sigma_1}\sigma_1(Z_{\mathbf{r}}^{\bar{\sigma}\sigma_2}Z_{\mathbf{r}+\mathbf{g}}^{\sigma_2\bar{\sigma}_1}-Z_{\mathbf{r}+\mathbf{g}}^{\bar{\sigma}\sigma_2}Z_{\mathbf{r}}^{\sigma_2\bar{\sigma}_1})C_{\mathbf{r}+\mathbf{a},\sigma_1}\Bigg), \tag{10}$$

$$\Omega_{\mathbf{k}}=\langle\{\mathscr{R}_{\mathbf{k},\sigma},\mathscr{B}^{\dagger}_{\mathbf{k},\sigma}\}\rangle=(\tau Q_{\tau}+JQ_J)/K_{\mathbf{k}}, \tag{11}$$

$$Q_{\tau}(\mathbf{k})=-\frac{7}{2}-8\left(\frac{1}{4}+C_{\mathbf{g}}\right)\gamma_{\mathbf{k}}+\left(\frac{1}{8}-C_{\mathbf{g}}+\frac{1}{2}\,C_{2\mathbf{g}}\right)\gamma_{2\mathbf{k}}+2\left(\frac{1}{8}-C_{\mathbf{g}}+\frac{1}{2}\,C_{\mathbf{d}}\right)\gamma_{\mathbf{dk}},$$

$$Q_J(\mathbf{k})=C_{\mathbf{g}}(\gamma_{\mathbf{k}}-4).$$

Here and below $\mathbf{d}=\mathbf{g}_x+\mathbf{g}_y$, and $\gamma_{\mathbf{dk}}=\cos(k_xg)\cos(k_yg)$.

Note that the expression (6) is formally exact. However, in contrast with the Dyson equation for the causal Green's functions, the diagrammatic representation is absent for the self-energy part (7). We see from Eqs. (6) and (7) that the self-energy $\Sigma(\mathbf{k},\omega)$, which accounts for the interaction effects, is expressed in terms of the higher-order Green's functions. One should notice, first, that the terms linear in $\mathscr{B}_{\mathbf{k},\sigma}$ do not contribute to the irreducible Green's function (7). Second, the lowest-order self-energy contribution is provided by the first term on the right-hand side of expression (7), while the second term leads to higher-order corrections. Following Ref. 7, we evaluate (7) with a proper decoupling procedure for the two-time correlation function $\langle\langle\mathscr{R}_{\mathbf{k},\sigma}(t)|\mathscr{R}^{\dagger}_{\mathbf{k},\sigma}(t')\rangle\rangle$. This procedure is equivalent to SCBA in a usual diagrammatic technique.[7] In our case this means that the two-time correlation function is decoupled into the spin–spin correlation function and the polaron–polaron correlation function. The adopted decoupling procedure preserves the main character of the polaron site operator (4)—four hole site operators surround the copper spin operator. It can be represented schematically in the form

$$\Bigg\langle Z_{\mathbf{r}_1}(t)\Bigg(\sum_{\mathbf{a}_1}c_{\mathbf{r}_2+\mathbf{a}_1}(t)Z_{\mathbf{r}_2}(t)\Bigg)\times\Bigg(\sum_{\mathbf{a}_2}Z_{\mathbf{r}_3}(t')c^{\dagger}_{\mathbf{r}_3+\mathbf{a}_2}(t')\Bigg)Z_{\mathbf{r}_4}(t')\Bigg\rangle\simeq\Bigg\langle\Bigg(\sum_{\mathbf{a}_1}c_{\mathbf{r}_2+\mathbf{a}_1}(t)Z_{\mathbf{r}_2}(t)\Bigg)\Bigg(\sum_{\mathbf{a}_2}Z_{\mathbf{r}_3}(t')c^{\dagger}_{\mathbf{r}_3+\mathbf{a}_2}(t')\Bigg)\Bigg\rangle\times\langle Z_{\mathbf{r}_1}(t)Z_{\mathbf{r}_4}(t')\rangle. \tag{12}$$

We note that the more complex decoupling procedure was also tested by us; it did not qualitatively alter the results given by approximation (12). In the next step we project the polaron operators in (12) onto $\mathscr{B}_{\mathbf{k}\sigma}$:

$$c_i(t)Z_j(t)\simeq\xi\mathscr{B}_{\mathbf{k}\sigma}(t),\quad \xi=\langle\{c_i(t)Z_j(t),\mathscr{B}^{\dagger}_{\mathbf{k},\sigma}\}\rangle. \tag{13}$$

Since we calculate only the irreducible part of the Green's function (7), the averages $\langle Z_{r_1}(t)Z_{\mathbf{r}_1}(t')\rangle$ are transformed to the corresponding spin-spin correlation functions $\langle S^{\alpha}_{\mathbf{r}_1}(t)S^{\alpha}_{\mathbf{r}_4}(t')\rangle$. Collecting all the terms, we have

$$\langle\langle\mathscr{R}_{\mathbf{k},\sigma}(t)|\mathscr{R}^{\dagger}_{\mathbf{k},\sigma}(t')\rangle\rangle\simeq N^{-1}\sum_{\mathbf{q}}\frac{K_{\mathbf{k}-\mathbf{q}}}{K_{\mathbf{k}}}\,\Gamma^2(\mathbf{k},\mathbf{q})\times\langle\langle\mathscr{B}_{\mathbf{k}-\mathbf{q},\sigma}(t)|\mathscr{B}^{\dagger}_{\mathbf{k}-\mathbf{q},\sigma}(t')\rangle\rangle\times\langle\mathbf{S}_{-\mathbf{q}}(t)\mathbf{S}_{\mathbf{q}}(t')\rangle, \tag{14}$$

where

$$\Gamma(\mathbf{k},\mathbf{q})=\tau\Gamma_{\tau}(\mathbf{k},\mathbf{q})+\frac{J}{2}\,\Gamma_J(\mathbf{k},\mathbf{q}),$$

$$\Gamma_{\tau}(\mathbf{k},\mathbf{q})=4\gamma_{\mathbf{k}-\mathbf{q}}\left(\frac{1+\gamma_{\mathbf{k}-\mathbf{q}}}{2K_{\mathbf{k}-\mathbf{q}}}-1\right),$$

$$\Gamma_J(\mathbf{k},\mathbf{q})=4\gamma_{\mathbf{q}}\left[\left(\frac{3}{4}-C_{\mathbf{g}}\right)\frac{4\gamma_{\mathbf{k}-\mathbf{q}}}{3K_{\mathbf{k}-\mathbf{q}}}-1\right],$$

$$\langle \mathbf{S}_{-\mathbf{q}}(t)\mathbf{S}_{\mathbf{q}}(t')\rangle = \frac{1}{N}\sum_{\mathbf{r},\mathbf{r}'} e^{i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})}\langle \mathbf{S}_{\mathbf{r}}(t)\mathbf{S}_{\mathbf{r}'}(t')\rangle.$$

Using the spectral representation for the Green's functions, we obtain the following intermediate result for the self-energy:

$$\Sigma(\mathbf{k},\omega) = \frac{1}{N}\sum_{\mathbf{q}} \frac{K_{\mathbf{k}-\mathbf{q}}}{K_{\mathbf{k}}}\Gamma^2(\mathbf{k},\mathbf{q}) \times \int_{-\infty}^{\infty}\frac{d\omega_1}{\pi}\int_{-\infty}^{\infty}\frac{d\omega_2}{\pi}\frac{e^{\beta(\omega_1+\omega_2)}+1}{(e^{\beta\omega_1}+1)(e^{\beta\omega_2}-1)} \times \frac{\mathrm{Im}[G(\mathbf{k}-\mathbf{q},\omega_1+i0^+)]\mathrm{Im}[D(\mathbf{q},\omega_2+i0^+)]}{\omega-(\omega_1+\omega_2)+i0^+}. \tag{15}$$

The spin excitation Green's function is[22,23]

$$D(\mathbf{q},\omega) = \langle S^z_{-\mathbf{q}}|S^z_{\mathbf{q}}\rangle = -\frac{8JC_{\mathbf{g}}}{3}\frac{1-\gamma_{\mathbf{q}}}{\omega^2-=\omega^2_{\mathbf{q}}}, \tag{16}$$

where

$$\omega^2_{\mathbf{q}} = -32J\alpha_1 C_{\mathbf{g}}(1-\gamma_{\mathbf{q}})(2\Delta+1+\gamma_{\mathbf{q}}).$$

We ignore the influence of doped holes on the copper spin dynamics and use the spin spectrum parameters calculated in Ref. 22 (the vertex correction $\alpha_1=1.7$, the spin excitations condensation part $m^2=0.0225$, and $\Delta=0$ at $T=0$).

As a result, we obtain the integral equation for the Green's function that always arises in the framework of the SCBA:

$$G(\mathbf{k},\omega) = \frac{1}{\omega-\Omega_{\mathbf{k}}-\Sigma(\mathbf{k},\omega)}, \tag{17}$$

where

$$\Sigma(\mathbf{k},\omega) = \frac{1}{N}\sum_{\mathbf{q}} M^2(\mathbf{k},\mathbf{q})[(1+\nu_{\mathbf{q}})G(\mathbf{k}-\mathbf{q},\omega-\omega_{\mathbf{q}}) + \nu_{\mathbf{q}}G(\mathbf{k}-\mathbf{q},\omega+\omega_{\mathbf{q}})]. \tag{18}$$

Here $\nu_{\mathbf{q}}=1/[\exp(\beta\omega_{\mathbf{q}})-1]$ is the Bose function and

$$M^2(\mathbf{k},\mathbf{q}) = \frac{K_{\mathbf{k}-\mathbf{q}}}{K_{\mathbf{k}}}\Gamma^2(\mathbf{k},\mathbf{q})\frac{(-4C_{\mathbf{g}})(1-\gamma_{\mathbf{q}})}{\omega_{\mathbf{q}}}. \tag{19}$$

The function $\Gamma(\mathbf{k},\mathbf{q})$ corresponds to the bare vertex for the coupling between a spin polaron and a spin wave. It is known[27] that this vertex is substantially renormalized for $\mathbf{q}$ close to the AFM vector $\mathbf{q}_0=(\pi,\pi)$. This renormalization is due to the strong interaction of a polaron with the condensation part of spin excitations that must be taken into account at the outset. As a result, the renormalized vertex $\tilde{\Gamma}(\mathbf{k},\mathbf{q})$ must be proportional to[27] $\sqrt{(\mathbf{q}-\mathbf{q}_0)^2+L_s^{-2}}$, where L_s is the spin–spin correlation length; $L_s\to\infty$ in the case of a long-range-order state of the spin subsystem. Below we take this renormalization into account empirically by the substitution

$$\Gamma(\mathbf{k},\mathbf{q})\to\tilde{\Gamma}(\mathbf{k},\mathbf{q})=\Gamma(\mathbf{k},\mathbf{q})\sqrt{1+\gamma_{\mathbf{q}}}. \tag{20}$$

The introduced vertex correction is proportional to $|\mathbf{q}-\mathbf{q}_0|$ for $\mathbf{q}$ close to $\mathbf{q}_0$. We have used also the following two functions for the vertex correction, $\sqrt{1+\gamma^3_{\mathbf{q}}}$ and $\sqrt{1+\gamma^5_{\mathbf{q}}}$, and have obtained the results similar to those presented below. Note that the bare vertex leads to a dramatic decrease of the quasiparticle band width.

3. SOLUTION OF THE INTEGRAL EQUATION

The equation (17) is usually solved by an iteration procedure. We propose here an alternative method, which is based on the continuous-fraction expansion of $G(\mathbf{k},z)$:

$$G(\mathbf{k},z) = \frac{b_0^2}{z-a_0-}\frac{b_1^2}{z-a_1-}\cdots\frac{b_n^2}{z-a_n-}\cdots,$$
$$a_n=a_n(\mathbf{k}),\quad b_n=b_n(\mathbf{k}), \tag{21}$$

where

$$b_0^2=\int_{-\infty}^{\infty}A_p(\mathbf{k},\omega)d\omega=K_{\mathbf{k}},$$
$$a_0=\frac{1}{b_0^2}\int_{-\infty}^{\infty}\omega A_p(\mathbf{k},\omega)d\omega=\Omega_{\mathbf{k}}.$$

The coefficients b_n, a_n, $n>0$ are related to the spectral density $A_p(\mathbf{k},\omega)$ via the set of orthogonal polynomials $P_n(\omega)$, which satisfy the recurrence:[28–32]

$$P_{-1}(\omega)=0,\quad P_0(\omega)=1,$$
$$P_{n+1}(\omega)=(\omega-a_n)P_n(\omega)-b_n^2P_{n-1}(\omega), \tag{22}$$

and

$$a_n=\frac{\int_{-\infty}^{\infty}\omega P_n^2(\omega)A_p(\mathbf{k},\omega)d\omega}{\int_{-\infty}^{\infty}P_n^2(\omega)A_p(\mathbf{k},\omega)d\omega}, \tag{23}$$

$$b_{n+1}^2=\frac{\int_{-\infty}^{\infty}P_{n+1}^2(\omega)A_p(\mathbf{k},\omega)d\omega}{\int_{-\infty}^{\infty}P_n^2(\omega)A_p(\mathbf{k},\omega)d\omega}. \tag{24}$$

Here we have used the nonnormalized form of the polynomials

$$\int_{-\infty}^{\infty}P_n(\omega)P_s(\omega)A_p(\mathbf{k},\omega)d\omega=\delta_{ns}\left(\prod_{m=1}^{m=n}b_m\right)^2.$$

Comparing Eqs. (21) and (17), we see that the self-energy $\Sigma(\mathbf{k},z)$ is the continuous fraction which is similar to $G(\mathbf{k},z)$. Thus we can introduce the spectral density

$$\rho(\mathbf{k},\omega)=-\mathrm{Im}[\Sigma(\mathbf{k},\omega+i0^+)]/\pi$$

and the set of polynomials $\Pi_n(\omega)$ with the recurrence analogous to (22):

$$\Pi_n(\omega)=(\omega-a_n)\Pi_{n-1}(\omega)-b_n^2\Pi_{n-2}(\omega),$$
$$\Pi_0(\omega)=1,\quad \Pi_{-1}(\omega)=0,$$

where

$$b_1^2=\int_{-\infty}^{\infty}\rho(\mathbf{k},\omega)d\omega,\quad a_1=\frac{1}{b_1^2}\int_{-\infty}^{\infty}\omega\rho(\mathbf{k},\omega)d\omega,$$
$$\alpha_{n+1}=\frac{\int_{-\infty}^{\infty}\omega\Pi_n^2(\omega)\rho(\mathbf{k},\omega)d\omega}{\int_{-\infty}^{\infty}\Pi_n^2(\omega)\rho(\mathbf{k},\omega)d\omega},$$

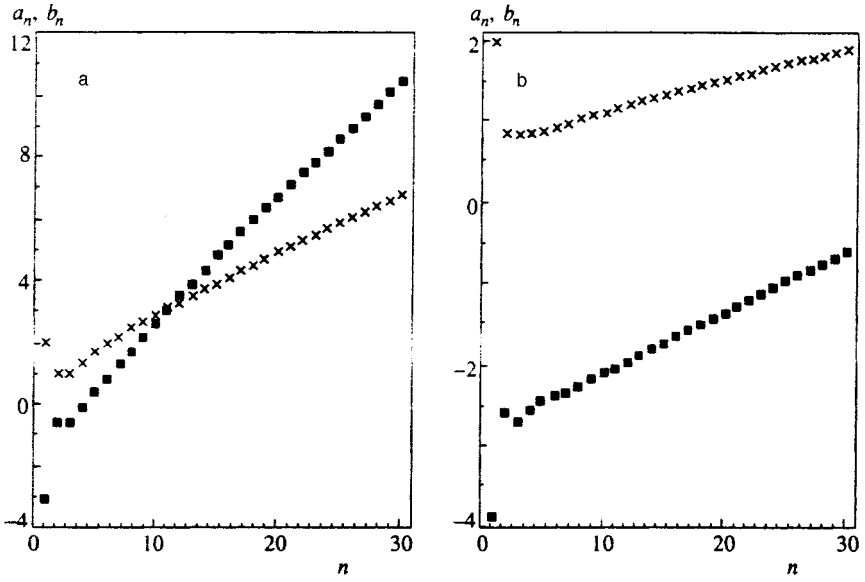

FIG. 1. The coefficients a_n (squares) and b_n (crosses) of the continuous-fraction expansion of $G_p(\mathbf{k},\omega)$ as functions on n for $\mathbf{k}=(\pi/2,\pi/2)$: a) $J=0.7\tau$; b) $J=0.1\tau$. Calculated on the 32×32 cell lattice.

$$b_{n+1}^2=\frac{\int_{-\infty}^{\infty}\Pi_n^2(\omega)\rho(\mathbf{k},\omega)d\omega}{\int_{-\infty}^{\infty}\Pi_{n-1}^2(\omega)\rho(\mathbf{k},\omega)d\omega},\quad n\geqslant 1. \tag{25}$$

On the other hand, we have the following relation from Eq. (18):

$$\rho(\mathbf{k},\omega)=\frac{1}{K_{\mathbf{k}}}\frac{1}{N}\sum_{\mathbf{q}} M^2(\mathbf{k},\mathbf{q})[(1+\nu_{\mathbf{q}})A_p(\mathbf{k}-\mathbf{q},z-\omega_{\mathbf{q}}) + \nu_{\mathbf{q}}A_p(\mathbf{k}-\mathbf{q},z+\omega_{\mathbf{q}})]. \tag{26}$$

Inserting the expression for $\rho(\mathbf{k},\omega)$ in Eq. (25), we can express the coefficients a_{n+1} and b_{n+2} in terms of the integrals of the form

$$\int_{-\infty}^{\infty}(\omega\pm\omega_{\mathbf{q}})^m\Pi_i^2A_p(\mathbf{k}-\mathbf{q},\omega)d\omega,\quad i\leqslant n,\quad m=0,1. \tag{27}$$

Now, the trick is that the polynomials in ω in the integrals (27) have the degree less than or equal to $2n+1$. As was proved by Nex,[29] such integrals may be expressed in terms of the coefficients $\{a_0,\ldots,a_n,b_0\ldots,b_n\}$. The details of such a procedure are presented in the Appendix. It turns out, therefore, that in SCBA we can recursively calculate pairs of coefficients a_{n+1},b_{n+1} and obtain $\Sigma(\mathbf{k},\omega)$ in the continuous-fraction form. Of course, we must calculate simultaneously the coefficients at all the chosen $\mathbf{k}+\mathbf{q}$ points in the first Brillouin zone. Below the chosen points correspond to a lattice of 32×32 unit cells. Our procedure allows us to avoid the iterative solution of Eq. (17) for complex energies.

4. TERMINATION OF THE CONTINUOUS FRACTION

The procedure outlined in the previous section would be efficient if after calculating a finite number of coefficients a_n, b_n, $n\leqslant n_0$, we could appropriately approximate that part (infinite in our case) of the continuous fraction T_{n_0} which has not been calculated. In other words, we rewrite the expression (21) in the form

$$G(\mathbf{k},z)=\frac{b_0^2}{z-a_0-}\frac{b_1^2}{z-a_1-}\cdots\frac{b_{n_0}^2}{z-a_{n_0}-T_{n_0}(\mathbf{k},z)} \tag{28}$$

and try to find a function $\tilde{T}_{n_0}$ (so-called «terminator») that is close to T_{n_0}.

Various ways to construct such approximations are described in the literature on the recursion method.[29–31] The asymptotic behavior of continuous-fraction coefficients is governed by the band structure and singularities of spectral density.[31] The main asymptotic behavior depends on the band structure: $\{a_n\}$ and $\{b_n\}$ converge toward limits in the single band case and oscillate endlessly in a predictable way in the multiband case. Damped oscillations are created by isolated singularities. The main point here is that an isolated simple pole produces exponentially damped contribution in $\{a_n\}$, $\{b_n\}$, $n\to\infty$. For our case it means that the quasiparticle pole position and weight could be obtained with high accuracy from finite number of coefficients, and the asymptotic behavior determines the incoherent part of the spectrum. It is obvious that the spectrum we deal with has a lower bound and no upper bound. We can thus expect that coefficients will not converge to some finite values but will rather tend to infinity.

In Fig. 1 we represent the coefficients a_n and b_n as functions of n calculated according to the procedure described in the preceding section for two values of J ($J=0.7\tau$ and $J=0.1\tau$) and for $\mathbf{k}=(\pi/2,\pi/2)$. We see that the distinctive feature of this dependence is that for large n the coefficients a_n and b_n are linear functions of n. Accordingly, the slope for a_n coefficients is twice as large as the slope for b_n. The behavior of the coefficients may therefore be approximated as

$$b_n\approx\lambda_1 n+\lambda_2,\quad a_n\approx2\lambda_1 n+\lambda_3,\quad \lambda_i=\lambda_i(\mathbf{k}),\quad n\gg1. \tag{29}$$

It is interesting that the coefficients for the $t-J$ model have analogous behavior when the slave-fermion Hamiltonian of this model[2] is treated in SCBA.[4] For $J=0.4t$, $\mathbf{k}$

$=(\pi/2,\pi/2)$ the coefficients a_n and b_n, which are governed by the relation analogous to (18), are shown in Fig. 2a.

Now we shall show that the same asymptotic expression (29) has the continuous-fraction expansion of the incomplete gamma function which is written as[33]

$$\Gamma(\alpha,x)=\frac{e^{-x}x^{\alpha}}{x+}\,\frac{1-\alpha}{1+}\,\frac{1}{x+}\,\frac{2-\alpha}{1+}\cdots \tag{30}$$

We shall use this circumstance for the construction of the terminator $\tilde{T}_N(\mathbf{k},z)$ for $G(\mathbf{k},z)$ [Eq. (28)].

We introduce the function

$$\tilde{g}(\alpha,x)=-\frac{\Gamma(\alpha,-x)}{e^{x}(-x)^{\alpha}}=\left(x-\frac{1-\alpha}{1-\theta_1}\right)^{-1}, \tag{31}$$

where

$$\theta_n=n\left(x-\frac{n+1-\alpha}{1-\theta_{n+1}}\right)^{-1}. \tag{32}$$

In order to rewrite the continuous fraction (31) in the form analogous to Eq. (21), we denote

$$\frac{1}{1-\theta_n}\equiv 1+nt_n. \tag{33}$$

We can then obtain the relations

$$t_n=\frac{1}{x-(2n+1-\alpha)-(n+1)(n+1-\alpha)t_{n+1}}, \tag{34}$$

so that

$$\tilde{g}(\alpha,x)=t_0 \tag{35}$$

has the form (28) with the coefficients $\tilde{b}_0^2=1$ and

$$\overline{a_n}=2n+1-\alpha,\quad \overline{b_n^2}=n(n-\alpha). \tag{36}$$

Comparing Eqs. (29) and (36) for large n, when $\sqrt{n(n-\alpha)}\approx n-\alpha/2$, we see that the substitution

$$\alpha=-\frac{2\lambda_2}{\lambda_1},\quad x=\frac{z+2\lambda_2-\lambda_3+\lambda_1}{\lambda_1}$$

leads to the function $\tilde{G}$:

$$\tilde{G}(\mathbf{k},z)=\frac{1}{\lambda_1}\,\tilde{g}\left(-\frac{2\lambda_2}{\lambda_1},\,\frac{z+2\lambda_2-\lambda_3+\lambda_1}{\lambda_1}\right), \tag{37}$$

which has the same asymptotic behavior as $G(\mathbf{k},z)$ [Eq. (21)]. This means that $\tilde{G}(\mathbf{k},z)$ can be used as the terminator for $G(\mathbf{k},z)$; i.e., we can express $\tilde{T}_{n_0}(\mathbf{k},z)=\overline{b}_{n_0+1}t_{n_0+1}$ in terms of $\tilde{G}(\mathbf{k},z)$ and the coefficients $\tilde{a}_n$, b_n, $n\leqslant n_0$, and then substitute it for $T_{n_0}(\mathbf{k},z)$ (see Ref. 30 for the particular features of the matching Greenians).

We thus obtain $G(\mathbf{k},z)$ in the total complex energy plane including the real axis. Note that usually the procedure of discretizing the energy range ω is used for the iteration process where the Dyson equation is solved numerically. It is not obvious that such a self-consistent solution leads to the correct analytical properties of the resulting Green's function. In contrast, the continuous-fraction representation guarantees these properties [e.g., the positive definiteness of the spectral function $A_p(\mathbf{k},\omega)$].

5. RESULTS AND DISCUSSION

In this section we present our results for the retarded Green's function $G(\mathbf{k},\omega)$ for the three-band model at $T=0$. The self-consistent equation (17) was solved on a 32×32 cell lattice. The number of calculated continuous-fraction levels n_0 was assumed to be $n_0=30$.

First, we check the validity of the method outlined above by calculating the spinless hole Green's function for the $t-J$ model and compare the results with the results of Martinez and Horsch[4] obtained by the usual iteration procedure. In Fig. 2 $A_h(\mathbf{k}_1,\omega+i\eta)$, $\mathrm{Re}\,\Sigma(\mathbf{k}_1,\omega)$, $-\mathrm{Im}\,\Sigma(\mathbf{k}_1,\omega)$, $\mathbf{k}_1=(\pi/2,\pi/2)$ for the value of $J=0.4t$ are represented for the 16×16 site lattice and broadening constant $\eta=0.01t$. Comparison of Fig. 2b–2d and the corresponding functions given in Figs. 7 and 8 from Ref. 4 (the same lattice size and the same η) demonstrates that the positions of the peaks of the hole spectral function and the peak's intensities coincide. The difference is that our $A_h(\mathbf{k}_1,\omega)$ is smoother and there are no strong oscillations in the self-energy $\Sigma(\mathbf{k},\omega)$ in the interval $-2t<\omega<-0.75t$.

The results for the small spin polaron spectral density, real and imaginary parts of the self-energy for the characteristic value of the energy parameter $J=0.7\tau$, are given in Fig. 3 for the symmetrical points $\mathbf{k}_1=(\pi/2,\pi/2)$, $\mathbf{k}_2=(0,0)$, $\mathbf{k}_3=(\pi,\pi)$. The energy broadening parameter is $\eta=0.002$ (we will refer to all quantities in units of τ from now on). The main common feature in the spectral density for $\mathbf{k}_1$ and $\mathbf{k}_2$ is the existence of a sharp quasiparticle peak at the bottom of each spectrum. The position of the quasiparticle peak corresponds to the condition $\mathrm{Re}\,G^{-1}(\mathbf{k},\omega)=0$, i.e., the point where we have the crossing of the functions $y=\omega-\Omega_{\mathbf{k}}$ and $\mathrm{Re}\,\Sigma(\mathbf{k},\omega)$, see Figs. 3a and 3b. In Fig. 3d we show $A_p(\mathbf{k}_1,\omega)$ calculated for $\eta=0.002$ (solid line) and $\eta=0.0005$ (dashed line) in order to study the scaling behavior of the peaks and their widths with respect to changes in η. Both peaks fit quite closely with a Lorentzian $(1/\pi)\times\{Z(\mathbf{k}_1)\,\eta/[(\omega-\epsilon(\mathbf{k}_1))^2+\eta^2]\}$, where $\epsilon(\mathbf{k}_1)$ is the location of the peak, which in the limit $\eta\to 0$ becomes $Z(\mathbf{k}_1)\,\delta(\omega-\epsilon(\mathbf{k}_1))$. This means that $\mathrm{Im}\,\Sigma(\mathbf{k},\omega_p)\to 0$ in the same limit. Here and below we speak about the position $\epsilon(\mathbf{k})$ of such peaks (with the imaginary part of the pole close to zero) in terms of the quasiparticle energy.

Figures 3a and 3b also demonstrate that the incoherent part of $A_p(\mathbf{k},\omega)$ increases and the pole strength decreases with increasing $\epsilon(\mathbf{k})$, $Z(\mathbf{k}_1)=0.82$, $Z(\mathbf{k}_2)=0.347$. We recall that $\Omega_{\mathbf{k}}$ represents the center of gravity of the spectral function. In our figures the center of gravity corresponds to the crossing of the real axis by the line $y=\omega-\Omega_{\mathbf{k}}$. Therefore, if the quasiparticle peak is far from this point, we would have a large incoherent part.

Quite different features are demonstrated in $A_p(\mathbf{k}_3,\omega)$ in Fig. 3c. The broad lowest peak is determined by the appearance of nonzero $\mathrm{Im}\,\Sigma(\mathbf{k}_3,\omega)$ in the region where $\mathrm{Re}\,G^{-1}(\mathbf{k}_3,\omega)$ has no zeros. Two broad additional peaks at $\omega\approx -1.6$ and $\omega\approx -1.05$ are formed because the zero values of $\mathrm{Re}\,G^{-1}(\mathbf{k}_3,\omega)$ are close to these ω. At the same time, the $\mathrm{Im}\,\Sigma(\mathbf{k}_3,\omega)$ is strong in these regions. Moreover, the maximum of $\mathrm{Im}\,\Sigma(\mathbf{k}_3,\omega)$ (near the point $\omega\approx -1.37$) determines

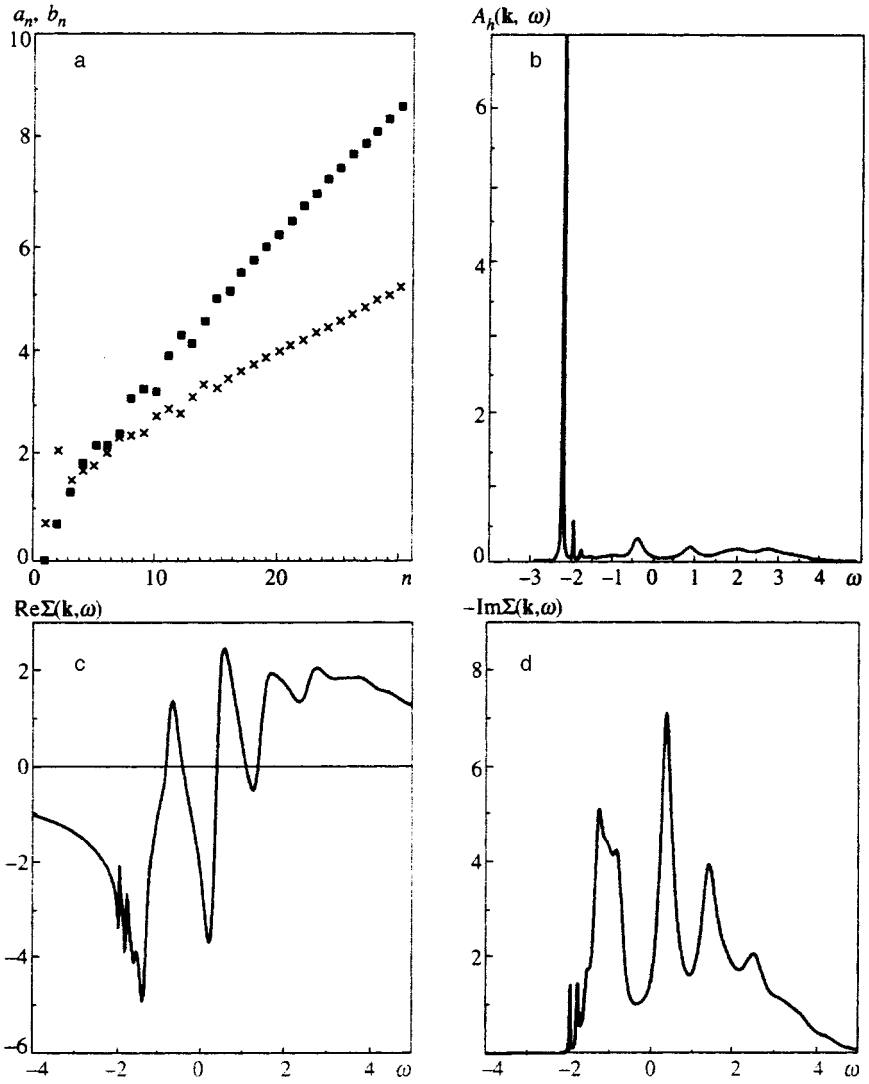

FIG. 2. Results for the hole Green's function $G_h(\mathbf{k},\omega)$ for the $t-J$ model calculated with the same parameters as in Ref. 4 ($J=0.4t$, $\mathbf{k}=(\pi/2,\pi/2)$, $\eta=0.01$, 16×16 site lattice): a) the coefficients a_n (squares) and b_n (crosses) of the continuous-fraction expansion of $G_h(\mathbf{k},\omega)$ as functions of n; b) spectral function $A_h(\mathbf{k},\omega)$; c) real part of the self-energy; d) imaginary part of the self-energy. The unit of energy is $t=1$.

the local minimum of $A_p(\mathbf{k}_3,\omega)$, despite of the fact that this point is close to the frequency where $\mathrm{Re}\,G^{-1}(\mathbf{k}_3,\omega)=0$. It is clear that it is impossible to treat any of the $A_p(\mathbf{k}_3,\omega)$ peaks as a quasiparticle peak. Bear in mind that the qualitative behavior of the real part of the self-energy in Fig. 3c is close to the one which is represented by Kampf and Schrieffer; see Fig. 3b in Ref. 34, for the pseudogap regime of the Hubbard model. Figure 3c gives three solutions of $\mathrm{Re}\,G^{-1}(\mathbf{k}_3,\omega)=0$. Although there is a sharp crossover from a situation with three solutions to one quasiparticle solution, the spectral function still changes smoothly due to the presence of the imaginary part of Σ.

Figures 3b and 3c demonstrate qualitatively a different character of $A_p(\mathbf{k},\omega)$ for the points $\mathbf{k}_2=(0,0)$, $\mathbf{k}_3=(\pi,\pi)$. This is a consequence of the spherically symmetric approach in treating the AFM copper spin subsystem. As mentioned in the Introduction, this approach gives rise to the spectral function periodicity relative to the full Brillouin zone, not the magnetic zone.

In Fig. 4 we show the dispersion relation $\epsilon(\mathbf{k})$ of the quasiparticle band and the mean field dispersion $\Omega_{\mathbf{k}}$ along the symmetry lines in the Brillouin zone. For $\epsilon(\mathbf{k})$ we reproduce only those $\mathbf{k}$ values for which the lowest peak has a pronounced quasiparticle peak, taking into account the following criteria: $-\mathrm{Im}\,\Sigma(\mathbf{k},\epsilon(\mathbf{k})+i\eta)<2\eta$. $\eta=0.002$. As we know,[15] due to the antiferromagnetic character of the spin correlation functions the $\Omega_{\mathbf{k}}$ demonstrates a «flat dispersion region» close to the line $\gamma_{\mathbf{k}}<0$, $|\gamma_{\mathbf{k}}|\leqslant1$, i.e., close to the boundary of the magnetic Brillouin zone $N-N-X$ (see Fig. 4). As we see from Fig. 4, the quasiparticle band exists in the greater part of the Brillouin zone except the region at the top of the $\Omega_{\mathbf{k}}$ spectrum. Moreover, the dispersion relation $\epsilon(\mathbf{k})$ qualitatively reproduces the main features of the spectrum $\Omega_{\mathbf{k}}$. As we mentioned in the Introduction, $\Omega_{\mathbf{k}}$ demonstrates the important features of the hole spectrum for CuO_2 plane if the O–O hoppings and spin frustration are taken into account.[21] We hope that in this case $\epsilon(\mathbf{k})$ will reproduce these features also.

Let us compare the small polaron spectral density $A_p(\mathbf{k},\omega)$ with the results for the bare hole $A_h(\mathbf{k},\omega)$ given by Kabanov and Vagov[12] for $\mathbf{k}_1=(\pi/2,\pi/2)$, $J=0.7\tau$ (see Fig. 3a). First, in Fig. 3a we see that $A_p(\mathbf{k}_1,\omega)$ has much sharper quasiparticle peaks relative to the results for a bare hole. For example, the pole strength $Z_p(\mathbf{k}_1)$ for the quasiparticle peak of $A_p(\mathbf{k},\omega)$ is $Z_p(\mathbf{k}_1)=0.82$. The corresponding value for A_h given by Ref. 12 is much smaller, $Z_h(\mathbf{k}_1)=0.25$.

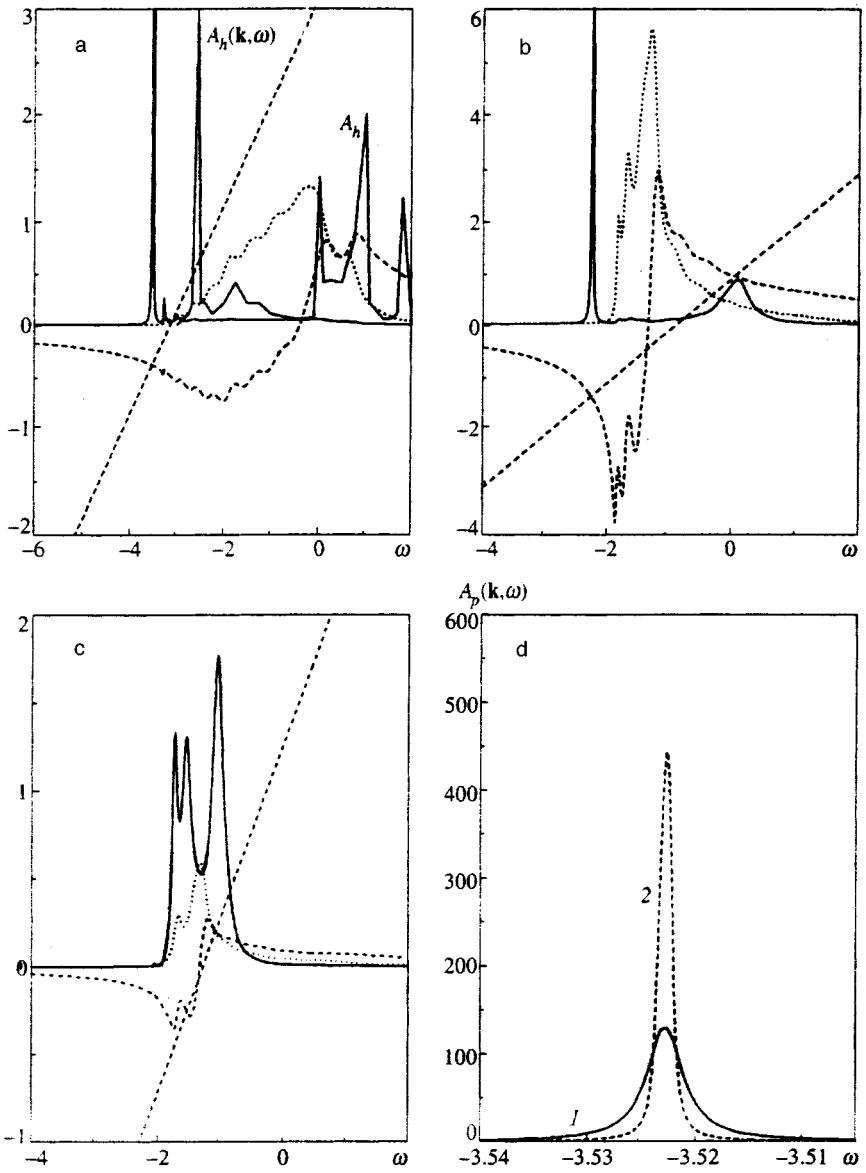

FIG. 3. Spin polaron spectral density ($A_p(\mathbf{k},\omega)$, solid lines), real ($\mathrm{Re}\,\Sigma(\mathbf{k},\omega)$, dashed lines) and imaginary ($-\mathrm{Im}\,\Sigma(\mathbf{k},\omega)$, dotted lines) parts of the self-energy calculated for $J=0.7\tau$, 32×32 cell lattice, and different $\mathbf{k}$: a) $\mathbf{k}=(\pi/2,\pi/2)$, here we also reproduce the hole spectral function $A_h(\mathbf{k},\omega)$, which was obtained in Ref. 12; b) $\mathbf{k}=(0,0)$; c) $\mathbf{k}=(\pi,\pi)$. In Fig. 1a–1c $\eta=0.002\tau$, the sloping straight lines represent the function $\omega-\Omega_{\mathbf{k}}$. d) The dependence of the quasiparticle peak of $A_p(\mathbf{k}=(\pi/2,\pi/2),\omega)$ for two values of the broadening factor η: 1— $\eta=0.02$; 2— $\eta=0.0004$. The unit of energy is $\tau=1$.

Second, Fig. 3a explicitly demonstrates the one-peak structure of $A_p(\mathbf{k}_1,\omega)$ in contrast to $A_h(\mathbf{k}_1,\omega)$. Finally, it is important that the bottom of the quasiparticle band $\epsilon(\mathbf{k})=-3.52$ is substantially lower than $\omega_h^{\min}=-2.6$ from Ref. 12. These results are a consequence of the fact that elementary excitation, i.e., spin polaron $\mathscr{B}_{\mathbf{k},\sigma}$, of small radii from the beginning involves the strong local hole-spin coupling.

It is clear that the quasiparticle peaks for a bare hole and a small polaron must coincide in the exact solution of the problem. The above-mentioned discrepancies between our calculations and those of Ref. 12 are the consequence of different approximations.

In order to test the convergence of our results relative to the increase of the lattice size and the number NL of calculated continuous-fraction levels, in Fig. 5 we show the quasiparticle peak of $A_p(\mathbf{k},\omega)$ at $\mathbf{k}=(\pi/2,\pi/2)$, $J=0.7$ for different lattices and n_0. This peak, as is evident from Figs. 5a and 5b, changes insignificantly in going from 24×24 to 32×32 cell lattice and from $NL=22$ to $NL=30$.

We consider now the transformation of $G(\mathbf{k},\omega)$ with the

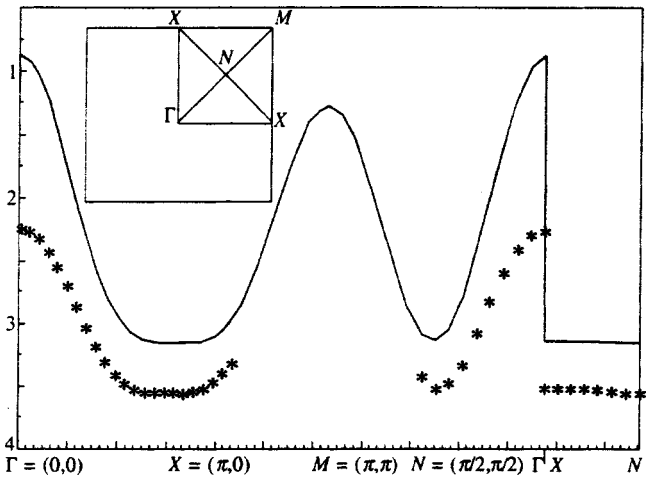

FIG. 4. The dispersion of the quasiparticle band $\epsilon(\mathbf{k})$ (symbols) and the mean field dispersion $\Omega_{\mathbf{k}}$ (solid line) along the symmetry lines in the Brillouin zone (see the inset) for $J=0.7$, 32×32 cell lattice, and $\eta=0.002$.

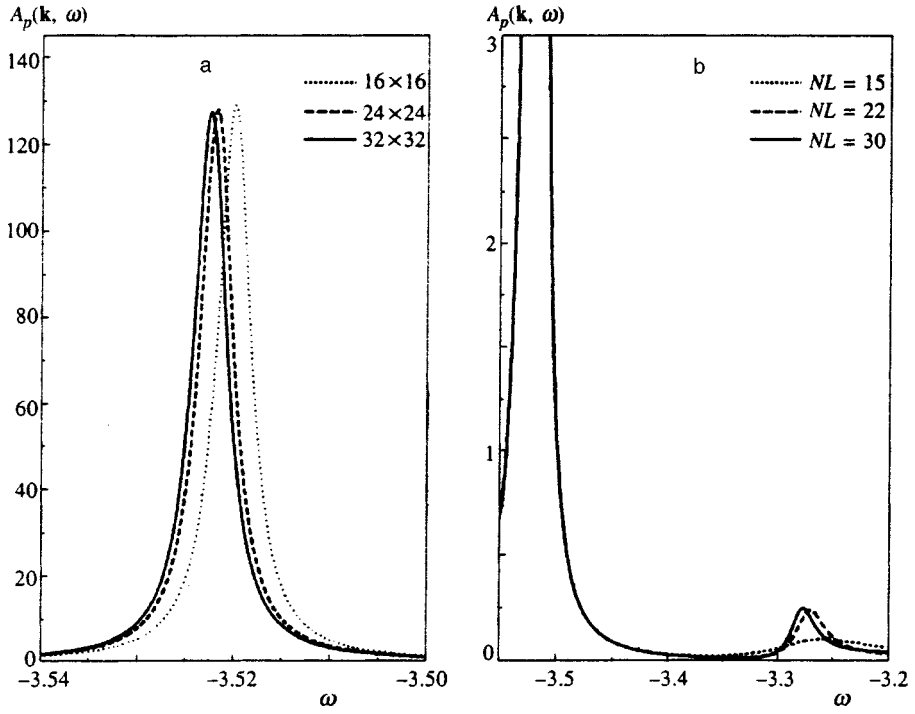

FIG. 5. The function $A_p(\mathbf{k}=(\pi/2,\pi/2),\omega)$ for $J=0.7$ and $\eta=0.002$ calculated for a) $NL=30$ and different lattice sizes; b) 32×32 cell lattice and different numbers NL of calculated continuous-fraction levels.

decrease of J. In order to clarify how the character of the $A_p(\mathbf{k},\omega)$ peaks is changed, in Fig. 6 we show $A_p(\mathbf{k},\omega)$ for the value of $J=0.1$ at points $\mathbf{k}_1=(\pi/2,\pi/2)$, $\mathbf{k}_2=(0,0)$, and $\mathbf{k}_3=(\pi,\pi)$. The decrease of J leads to the enlargement of the broad, incoherent part of $A_p(\mathbf{k},\omega)$.

As before, the flat band region of the quasiparticle band bottom enlarges along a magnetic Brillouin zone boundary. It is represented by the point $\mathbf{k}_1$. In Fig. 6a $A_p(\mathbf{k}_1,\omega)$ demonstrates explicitly a rather strong quasiparticle peak, $Z_p(\mathbf{k}_1)=0.5$, which corresponds to the condition $\mathrm{Re}\, G^{-1}(\mathbf{k},\omega)=0$. Quite different character of $A_p(\mathbf{k},\omega)$ is typical for k that correspond to the tops of $\Omega_{\mathbf{k}}$ band: In the low-energy sector for $\mathbf{k}_2,\mathbf{k}_3$ (see Figs. 6b and 6c) one observes low-intensity $A_p(\mathbf{k},\omega)$ peaks. For example, the pole strength Z_p of such a quasiparticle peak for $A_p(\mathbf{k}_2,\omega)$ is $Z_p(\mathbf{k}_2)=0.016$. Assuming $\omega_l(\mathbf{k})$ to be the value of ω corresponding to the center of these lowest-energy peaks, we see that $\mathrm{Re}\, G^{-1}(\mathbf{k},\omega_l(\mathbf{k}))\neq 0$ for the $\mathbf{k}$ under discussion. Figures 6b and 6c demonstrate that these peaks are determined by the peaks in $\mathrm{Im}\,\Sigma(\mathbf{k},\omega)$ at the points $\omega_l(\mathbf{k})$. The self-energy part $\Sigma(\mathbf{k},\omega)$ occurs through the Green's function of a small polaron bounded to spin waves. These peaks can be considered as the quasiparticle band of such complex states.

If we treat the width W of the quasiparticle band as the difference $\omega_l(\mathbf{k}_2=(0,0))-\omega_l(\mathbf{k}_1=(\pi/2,\pi/2))$, then W turns out to be of the order of J for small values of J ($J\simeq0.1$), consistent with the results for the hole Green's function approach.[12]

It is clear that for small J/τ the concept of a small spin polaron fails and it is important to estimate the validity limits of this concept. Our calculations demonstrate that the intensity of quasiparticle peaks and the structure of $A_p(\mathbf{k},\omega)$ do not change dramatically for $\mathbf{k}$, which corresponds to the band bottom, up to $J/\tau=0.1$. For example, $Z_p(\pi/2,\pi/2)\approx Z_p(\pi,0)\approx0.50$ at $J/\tau=0.1$. Therefore, the J/τ lowest boundary value of the small spin polaron concept validity is lower than $J/\tau=0.1$.

Table I presents the numerical values $\omega_l(\mathbf{k})$ of the center position of the lowest $A_p(\mathbf{k}_2,\omega)$ peaks [$\omega_l(\mathbf{k})=\epsilon(\mathbf{k})$ for $\mathbf{k}$ values where the quasiparticle peak is observed) and their pole strength (area under the peak) $Z_p(\mathbf{k})$ for $\mathbf{k}=(\pi/2,\pi/2),(0,0),(0,\pi)$ and different values of J.

We do not represent the results for large J ($J\gg\tau$) since our approach in the present form fails to describe this limit. Here, from the very beginning we treat a small polaron by a single site operator $\mathscr{B}_{\mathbf{r},\sigma}$ (4). For large J the mean-field static energy of such a state is proportional to J and such a state is unstable. In this limit, therefore, we must extend the basis of the site operators. The simplest way to do this is to include in the basis the additional operator of a bare hole. In SCBA this will lead to the system of two self-consistent equations. As a result, all effects of interaction between a spin subsystem and holes will be proportional to τ/J. The more general procedure for extending the small polaron operator basis is outlined in Ref. 21.

6. SUMMARY

We have studied the small spin polaron motion in the three-band model. The two-time retarded Green's function was calculated in the framework of self-consistent Born approximation for 32×32 cell lattice. We have shown that spin polaron of small radius represents a good approximation of the true quasiparticle low-energy excitation even at the mean-field level. Allowance for the self-energy does not crucially change the polaron motion picture for realistic values of parameters. For quasimomenta $\mathbf{k}$ values, which correspond to the band bottom, most of the total spectral weight is

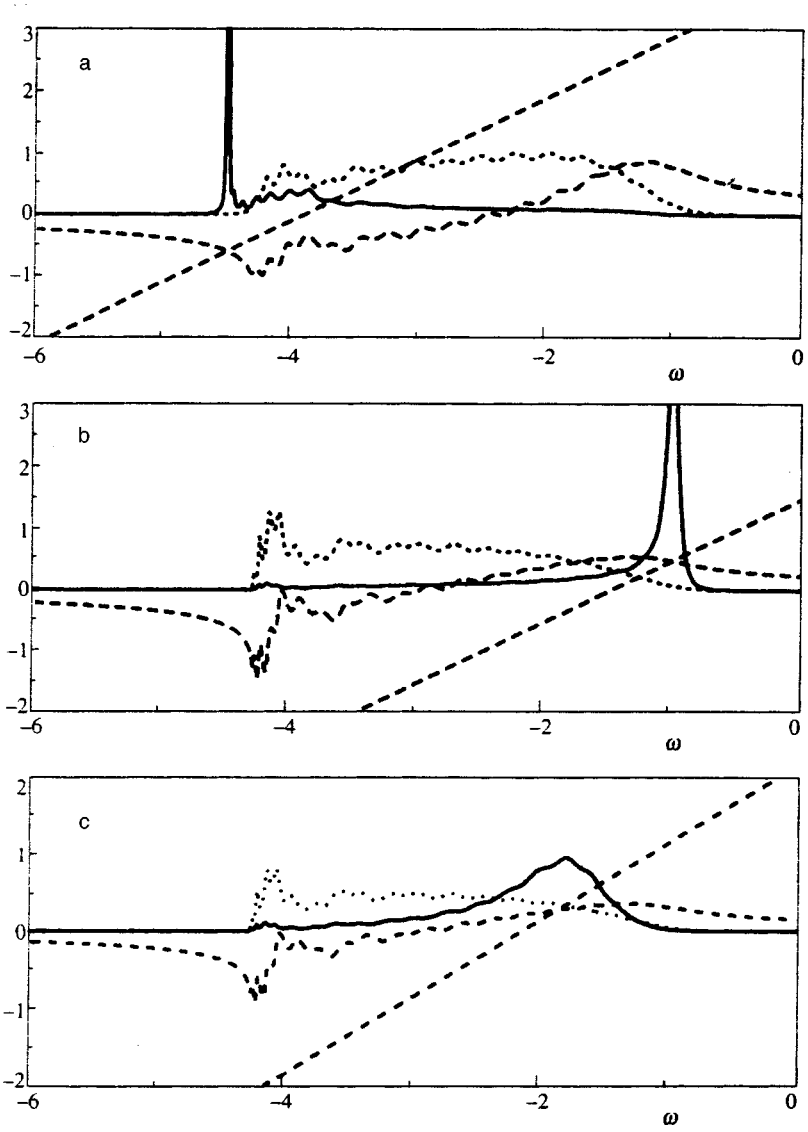

FIG. 6. Spin polaron spectral density ($A_p(\mathbf{k},\omega)$ solid lines), real ($\mathrm{Re}\,\Sigma(\mathbf{k},\omega)$, dashed lines) and imaginary ($-\mathrm{Im}\,\Sigma(\mathbf{k},\omega)$, dotted lines) parts of the self-energy $\Sigma(k,\omega)$ calculated for $J=0.1\tau$, $\eta=0.002$, 32×32 cell lattice at three different $\mathbf{k}$: a) $\mathbf{k}=(\pi/2,\pi/2)$; b) $\mathbf{k}=(0,0)$; c) $\mathbf{k}=(\pi,\pi)$.

concentrated in the quasiparticle peak (Table I). In the same region of $\mathbf{k}$-space the shape of the quasiparticle dispersion curve $\epsilon(\mathbf{k})$ reproduces the shape of the mean-field dispersion $\Omega_{\mathbf{k}}$ (Fig. 4).

We compared our results with the previous studies[11,12] which started from the bare hole. We see that the small polaron mean-field energy $\Omega_{\mathbf{k}}$ lies much lower than the quasiparticle pole obtained from SCBA for the bare hole. Since $\Omega_{\mathbf{k}}$ determines the center of gravity for the Green's function spectral density, the actual quasiparticle pole position (at least for the band bottom) should lie deeper in energy than $\Omega_{\mathbf{k}}$ (Figs. 3 and 6). This means that in the three-band model the important local correlations should be taken into account in zero approximation and small spin polaron should be constructed. The polaron scattering on spin waves will then be of less importance and it may be treated by perturbation methods.

We conclude that the low-energy physics of high-T_c superconductors should be considered in terms of small spin polaron dynamics. In particular, the problem of superconducting hole pairing must be treated as pairing of these quasiparticles rather than pairing of bare holes.

Finally, we wish to clarify the difference between our approach and those that use a Néel-type spin subsystem state. Our calculations are based on the approach developed in Refs. 22 and 23, where it is shown that a two-dimensional

TABLE I. Position $\omega_l(\mathbf{k})$ of the lowest in energy peak and the area $Z_p(\mathbf{k})$ under the peak for different values of J/τ and $\mathbf{k}$.

J/τ	$Z_p(0,0)$	$\omega_l(0,0)$	$Z_p(\pi/2,\pi/2)$	$\omega_l(\pi/2,\pi/2)$	$Z_p(\pi,0)$	$\omega_l(\pi,0)$
0.1	0.016	−4.24	0.50	−4.48	0.55	−4.51
0.3	0.039	−3.37	0.72	−4.09	0.714	−4.13
0.5	0.174	−2.714	0.793	−3.79	0.738	−3.83
0.7	0.347	−2.25	0.823	−3.52	0.808	−3.56

$s=1/2$ antiferromagnetic system at low temperature is in the rotationally invariant state that preserves this invariance in the limit where T goes to zero and the correlation length goes to infinity. Then, even for $T=0$, the points (0,0) and (π,π) are not equivalent due to the numerator in the spin Green's function [see Eq. (16)]. As a result, the short-range spin-spin correlation functions, such as $C^{+-}_{\mathbf{r},\mathbf{R}}=\langle S^{+}_{\mathbf{r}} S^{-}_{\mathbf{r}+\mathbf{R}}\rangle$, do not depend on the site $\mathbf{r}$. We note that according to Marshall's theorem,[35] the ground state of an antiferromagnet may be a spherically symmetric singlet state in the limit $N\to\infty$. The question about the ground state of a two-dimensional $s=1/2$ AFM system is not established exactly. Since at any finite temperature the rotationally invariant theory is true for the paramagnetic system with strong short-range AFM correlations, we think that it is impossible for the system to undergo an abrupt transition to a state with a spontaneously broken symmetry at $T=0$. Such a transition would mean an abrupt rise of the $\mathbf{r}$-dependence in $C^{+-}_{\mathbf{r},\mathbf{R}}$ and finite $\langle S^{z}_{\mathbf{r}}\rangle$. We believe that the ground state of a two-dimensional $s=1/2$ AFM system is a spherically symmetric state with a long-range order, and that it does not lead to the symmetry of the reduced antiferromagnetic zone. The energy bands therefore change infinitesimally as we go from $T=0$ to a temperature that is infinitesimally above zero.

In addition, our results demonstrate that the nonequivalence of the (0,0) and (π,π) points for the spectral function $A(\mathbf{k},\omega)$ and $\Omega_{\mathbf{k}}$ do not contradict the exact diagonalization studies for the ground state of the $t-J$ model (see, e.g., Figs. 18a and 23 in Ref. 1).

A few words about the scenario which assumes that a smooth transition from rotationally invariant state to the state with spontaneously broken symmetry takes place. Such a transition implies a smooth increase in $\langle S^{z}_{\mathbf{r}}\rangle$ and $\mathbf{r}$-dependence of $C^{+-}_{\mathbf{r},\mathbf{R}}$. The main effect consists not in the changes of $\Omega_{\mathbf{k}}$ and $A(\mathbf{k},\omega)$ at (0,0) and (π,π) points but in the opening of the gap along the boundary of the magnetic (reduced) Brillouin zone. As a result, we shall have two different bands with the periodicity of the magnetic Brillouin zone, i.e., the (0,0) and (π,π) points become equivalent for each of the two bands.

We are grateful to O. A. Starykh and P. Horsch for valuable discussions and comments. This work was supported, in part, by the INTAS-PFBR organization under project INTAS-RFBR No. 95-0591, by the Russian Fund for Fundamental Research (Grants Nos. 98-02-17187 and 98-02-16730), and by the Russian National Program on Superconductivity (Grant No 93080).

APPENDIX

Chain representation and integrals over the spectral density

The integration over spectral density could be done with the quadrature approach,[29] which is very efficient when applied to the electron structure calculations.[32] Unfortunately, the spectral density we deal with has no upper bound and depends exponentially on the energy; i.e., it substantially differs from the typical spectral density that appears in band-structure calculations. It turned out that the direct application of Nex's quadrature approach[29] is not stable numerically for our purpose. For calculations of the integrals (27) we therefore use the chain representation of the continuous fraction. This means that the continuous-fraction expansion of the form (21) may be interpreted as the Green's function $G(\omega)=\langle u_0|(\omega-\hat{h})^{-1}|u_0\rangle$ of the one-particle, tight-binding Hamiltonian $\hat{h}$ of the semi-infinite one-dimensional lattice with a_n, b_n, and $|u_n\rangle$ as the site energies, nearest-neighbor hoppings and on-site basis states, respectively:

$$a_n=\langle u_n|\hat{h}|u_n\rangle, \quad b_{n+1}=\langle u_n|\hat{h}|u_{n+1}\rangle.$$

We introduce the eigenstates $|\psi_m\rangle$ and eigenenergies E_m of the chain Hamiltonian:

$$\hat{h}=\sum_m |\psi_m\rangle E_m\langle\psi_m|.$$

The Green's function spectral density then becomes the local density of states at the zeroth site of the chain:[28]

$$A(\omega)=-\frac{1}{\pi}\,\mathrm{Im}\,G(\omega+i0^{+})$$

$$=\sum_m \langle u_0|\psi_m\rangle\delta(\omega-E_m)\langle\psi_m|u_0\rangle.$$

The following identities will then hold:

$$F=\int_{-\infty}^{\infty} f(\omega)A(\omega)d\omega=\int_{-\infty}^{\infty} f(\omega)\sum_m \langle u_0|\psi_m\rangle\delta(\omega$$

$$-E_m)\langle\psi_m|u_0\rangle d\omega=\sum_m \langle u_0|\psi_m\rangle f(E_m)\langle\psi_m|u_0\rangle$$

$$=\langle u_0|f(\hat{h})|u_0\rangle. \tag{38}$$

Nex[29] has proved that for a polynomial f of the degree $2n_0+1$ the integral F for the infinite chain has the same value as an analogous integral for the truncated chain of length n_0+1. The Hamiltonian of the truncated chain in the basis of the states $\{|u_0\rangle\ldots|u_{n_0}\rangle\}$ has the form of the tridiagonal $(n_0+1)\times(n_0+1)$ matrix:

$$h_T=\begin{bmatrix} a_0 & b_1 & & & \\ b_1 & a_1 & b_2 & & \\ & & \cdots & & \\ & & & a_{n_0-1} & b_{n_0} \\ & & & b_{n_0} & a_{n_0} \end{bmatrix}.$$

Now, instead of integrating the spectral density function over ω, we directly calculate the matrix function $f(h_T)$ in order to use f_{00} matrix element. Then $F=f_{00}$, as follows from the last identity in Eq. (38). We see that the result for F is expressed only in terms of the first coefficients $\{a_0,\ldots,a_n,b_0\ldots,b_n\}$.

[1] E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994).
[2] S. Schmitt-Rink, C. M. Varma, and A. E. Ruchenstein, Phys. Rev. Lett. **60**, 2793 (1988); F. Marsiglio, A. E. Ruchenstein, S. Schmitt-Rink, and C. M. Varma, Phys. Rev. B **43**, 10882 (1991).
[3] C. L. Kane, P. A. Lee, and N. Read, Phys. Rev. B **39**, 6880 (1989).
[4] G. Martinez and P. Horsch, Phys. Rev. B **44**, 317 (1991).

[5] Z. Lui and E. Manousakis, Phys. Rev. B **44**, 2414 (1991); **45**, 2425 (1992).
[6] A. Sherman and M. Schriber, Phys. Rev. B **48**, 7492 (1993); **50**, 12887 (1994); **50**, 6431 (1994).
[7] N. M. Plakida, V. S. Oudovenko, and V. Yu. Yushankhai, Phys. Rev. B **50**, 6431 (1994).
[8] A. Ramsak and P. Prelovsek, Phys. Rev. B **42**, 10415 (1990).
[9] V. J. Emery, Phys. Rev. Lett. **58**, 2794 (1988).
[10] V. J. Emery and G. Reiter, Phys. Rev. B **38**, 4547 (1988).
[11] O. F. de Alcantara Bonfim and G. F. Reiter, in *Proceedings of the Univ. of Miami Workshop on Electronic Structure and Mechanisms for High-Temperature Superconductivity*, ed. by J. Ashkenazi, Plenum Press, New York (1991).
[12] V. V. Kabanov and A. Vagov, Phys. Rev. B **47**, 12134 (1993).
[13] O. A. Starykh, O. F. de Alcantara Bonfim, and G. F. Reiter, Phys. Rev. B **52**, 12534 (1995).
[14] F. C. Zhang and T. M. Rice, Phys. Rev. Lett. **37**, 3759 (1988).
[15] A. F. Barabanov, L. A. Maksimov, and G. V. Uimin, JETP Lett. **47**, 622 (1988); Zh. Éksp. Teor. Fiz. **96**, 655 (1989) [Sov. Phys. JETP **69**, 371 (1989)]; A. F. Barabanov, R. O. Kuzian, and L. A. Maksimov, J. Phys. Cond Matter **3**, 9129 (1991).
[16] J. G. Tobin, C. G. Olson, C. Gu *et al.*, Phys. Rev. B **45**, 5563 (1992).
[17] K. Gofron, J. C. Campuzano, H. Ding *et al.*, J. Phys. Chem. Solids **54**, 1193 (1993).
[18] A. A. Abrikosov, J. C. Campuzano, and K. Gofron, Physica C **214**, 73 (1993).
[19] D. S. Dessau, Z.-X. Shen, D. M. King *et al.*, Phys. Rev. Lett. **71**, 2781 (1993).
[20] D. M. King, Z.-X. Shen, D. S. Dessau *et al.*, Phys. Rev. Lett. **73**, 3298 (1994).
[21] A. F. Barabanov, V. M. Berezovsky, E. Žasinas, and L. A. Maksimov, Zh. Éksp. Teor. Fiz. **110**, 1480 (1996) [JETP **83**, 819 (1996)].
[22] H. Shimahara and S. Takada, J. Phys. Soc. Jpn. **60**, 2394 (1991).
[23] A. Barabanov and O. Starykh, J. Phys. Soc. Jpn. **61**, 704 (1992); A. Barabanov and V. M. Berezovsky, Zh. Éksp. Teor. Fiz. **106**, 1156 (1994) [JETP **79**, 627 (1994)].
[24] A. F. Barabanov, R. O. Kuzian, and L. A. Maksimov, Phys. Rev. B **55**, 4015 (1997).
[25] H. Mori, Prog. Theor. Phys. **33**, 423 (1965); **34**, 399 (1965).
[26] N. M. Plakida, Phys. Lett. A **43**, 481 (1973).
[27] J. R. Schrieffer, J. Low Temp. Phys. **99**, 397 (1995).
[28] R. Haydock, in *Solid State Physics* **35**, ed. by H. Ehrenreich, F. Seitz, and D. Turmbull, Academic Press, New York (1980).
[29] C. M. M. Nex, J. Phys. A: Math Gen. **11**, 653 (1978).
[30] R. Haydock and C. M. M. Nex, J. Phys. C: Solid State Phys. **18**, 2235 (1985).
[31] *The Recursion Method and its Applications*, ed. by D. G. Pettifor and D. L. Weaire, Springer, Berlin (1985).
[32] R. Haydock and C. M. M. Nex, J. Phys. C: Solid State Phys. **17**, 4783 (1984).
[33] *Handbook of Mathematical Functions*, ed. by M. Abramowitz and I. A. Stegun, Nat. Bur. of Standards (1964).
[34] A. P. Kampf and J. R. Schrieffer, Phys. Rev. B **42**, 7967 (1990).
[35] W. Marshall, Proc. R. Soc. London, Ser. A **232**, 48 (1955); E. Lieb and D. C. Mattis, J. Math. Phys. **3**, 749 (1962).

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

Kinetics of heavy fermions

A. V. Gol'tsev*)

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted 8 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 1778–1786 (May 1998)

A transport equation is derived for the distribution function of heavy fermions in electric and magnetic fields that allows for potential and spin–spin interactions between the heavy fermions. The spectrum of the spin waves in the paramagnetic heavy-fermion state is calculated. Finally, processes associated with pair collisions of heavy fermions and the scattering of such fermions by charged impurities are studied. © *1998 American Institute of Physics.* [S1063-7761(98)01505-4]

1. INTRODUCTION

In heavy-fermion compounds at temperatures below the Kondo temperature T_0, the antiferromagnetic exchange interaction between conduction electrons and electrons localized in the partially filled f shell of rare-earth ions leads to a coherent Kondo effect. This manifests itself in the formation near the Fermi surface of quasiparticle states that have an effective mass that is two orders of magnitude greater than the electron mass in normal metals. Such quasiparticles became known as heavy fermions, and this name gave rise to an entire class of metals, which became known as heavy-fermion metals. Experiments have shown that at low temperatures, $T \ll T_0$, heavy-fermion compounds behave as normal Fermi liquids. This manifests itself in the nature of the temperature behavior of the resistance ($R = R_0 + AT^2$), the magnetic susceptibility ($\chi = \chi(0) + bT^2$), heat capacity ($C = \gamma T$), and other physical properties. Note that the parameters A, $\chi(0)$, b, and γ have anomalously large values in comparison to the values for normal metals, e.g., $\chi(T=0) \sim 10^2 \chi_0$, $\gamma \sim 10^2 \gamma_0$, and $A \sim 10^4 A_0$. According to the theory of normal Fermi liquids, $\chi(T=0)$ and γ are proportional to the effective mass m^* of quasiparticles on the Fermi surface. Hence the above experimental data suggest that heavy fermions have a mass m^* of order $10^2 m_0$. The presence on the Fermi surface of heavy quasiparticles was corroborated in studies of the de Haas–van Alphen effect. A comprehensive list of the experimental papers can be found in the reviews in Refs. 1 and 2. In a state with heavy fermions in some heavy-fermion compounds magnetic and superconducting phase transitions may occur into states with properties that differ considerably from those of normal metals.[1,2] Only some of these properties have been explained by the theory.

The microscopic theory of the heavy-fermion state is based on the Anderson lattice model (see, e.g., the review article in Ref. 3). However, in analyzing the magnetic properties, magnetic phase transitions, and transport phenomena in which the competition between the coherent Kondo effect and magnetism may play an important role, the microscopic approach encounters serious difficulties because it is still unknown how to account for the magnetic interaction between heavy fermions.

The Fermi-liquid nature of the heavy-fermion state suggests using Landau's phenomenological theory of Fermi liquids to describe the thermodynamic and transport properties. Such an approach was proposed in Refs. 4 and 5, and its main merit consists in the following. First, the approach makes it possible to take into account in a self-consistent manner the potential and exchange interactions between heavy fermions. Second, it provides a transparent physical interpretation of phenomena in heavy-fermion compounds.

This paper uses the Landau theory for normal Fermi liquids to develop the kinetic theory of heavy-fermion compounds. A transport equation is derived for the distribution function of heavy fermions in electric and magnetic fields that allows pair collisions of quasiparticles and scattering by charged impurities. The paper also studies the spectrum of spin waves.

2. TRANSPORT EQUATION FOR HEAVY FERMIONS

Phenomenological Fermi-liquid theory of the heavy-fermion state[4,5] is based on two assumptions: (1) the heavy-fermion wave function is a linear combination of the wave functions of the electron states in a broad conduction band and a narrow f-band, and (2) because of strong single-site Coulomb repulsion the concentration of f-electrons does not depend on time t and radius vector $\mathbf{r}$. The first assumption implies that the distribution of electrons over states with the wave vector $\mathbf{k}$ is described by a Hermitian matrix $N^{ab}_{\alpha\beta}(\mathbf{k},\mathbf{r},t)$, where the band indices a and b take values c and f, and α and β are spin indices for spin $\frac{1}{2}$. The diagonal elements $N^{cc}_{\alpha\beta}(\mathbf{k},\mathbf{r},t)$ and $N^{ff}_{\alpha\beta}(\mathbf{k},\mathbf{r},t)$ describe the distribution of electrons in the conduction band and the f-band. The off-diagonal elements characterize the formation of coherence between the conduction electrons and the f-electrons. To find $\hat{N}$, we must solve the following transport equation (the standard equation in the Landau theory of Fermi liquids):

$$\frac{\partial \hat{N}}{\partial t} + \{\nabla_{\mathbf{r}} \hat{N} \nabla_{\mathbf{k}} \hat{\varepsilon}\} - \{\nabla_{\mathbf{k}} \hat{N} \nabla_{\mathbf{r}} \hat{\varepsilon}\} - i[\hat{\varepsilon}, \hat{N}] = \hat{I}(\hat{N}), \qquad (1)$$

where $2\{\hat{A}\hat{B}\}=\hat{A}\hat{B}+\hat{B}\hat{A}$, $[\hat{A},\hat{B}]=\hat{A}\hat{B}-\hat{B}\hat{A}$, and $\hat{I}(\hat{N})$ is the collision integral. The quasiparticle energy matrix $\hat{\varepsilon}$ is a functional of $\hat{N}$ (Refs. 4 and 5). According to the second assumption, the transport equation (1) must be solved under the condition that the f-electron concentration is constant:

$$\sum_{\mathbf{k}\alpha} N^{f}_{\alpha\alpha}(\mathbf{k},\mathbf{r},t)=N_f. \tag{2}$$

Let us study the small deviations of the distribution matrix $\hat{N}=\hat{N}_0+\hat{N}_1$ (these deviations are characterized by a wave vector $\mathbf{q}$ and a frequency ω) from the equilibrium value $\hat{N}_0$:

$$\hat{N}_1(\mathbf{k},\mathbf{r},t)=\hat{N}_1(\mathbf{k},\mathbf{q},\omega)\exp(i\mathbf{q}\cdot\mathbf{r}-i\omega t)+\text{c.c.} \tag{3}$$

The physics of the problem implies that the fluctuations $\hat{N}_1$ are related to variations in the heavy-fermion distribution function. The relation can be found by using the relationship between $N^{ab}_{1\alpha\beta}(\mathbf{k},\mathbf{q},t)$ and the time evolution of the operator of an electron–positron pair with momentum $\mathbf{q}$:

$$\rho^{ab}_{\alpha\beta}(\mathbf{k},\mathbf{q})=b^{\dagger}_{\beta\mathbf{k}+\mathbf{q}/2}a_{\alpha\mathbf{k}-\mathbf{q}/2}, \tag{4}$$

where for $a,b=c,f$ the operators $c^{\dagger}_{\alpha\mathbf{k}}$, $f^{\dagger}_{\alpha\mathbf{k}}$ and $c_{\alpha\mathbf{k}}$, $f_{\alpha\mathbf{k}}$ are the operators of electron creation and annihilation in the conduction band and the f-band, respectively. Following Ref. 6, we assume that

$$N^{ab}_{1\alpha\beta}(\mathbf{k},\mathbf{q},t)=\langle\rho^{ab}_{\alpha\beta}(\mathbf{k},\mathbf{q},t)\rangle. \tag{5}$$

The operators $c_{\alpha\mathbf{k}}$ and $f_{\alpha\mathbf{k}}$ are related to the annihilation operators $g_{1\alpha\mathbf{k}}$ and $g_{2\alpha\mathbf{k}}$ of quasiparticle states in two hybrid bands $E_{1\mathbf{k}\alpha}$ and $E_{2\mathbf{k}\alpha}$ by a unitary transformation $\hat{U}_{\mathbf{k}}$:

$$\begin{pmatrix} g_{1\alpha\mathbf{k}} \\ g_{2\alpha\mathbf{k}} \end{pmatrix}=\hat{U}_{\mathbf{k}}\begin{pmatrix} c_{\alpha\mathbf{k}} \\ f_{\alpha\mathbf{k}} \end{pmatrix}. \tag{6}$$

Note that the matrix $\hat{U}_{\mathbf{k}}$ also reduces $\hat{N}_0$ to diagonal form, $\hat{N}_0=\hat{U}^{-1}_{\mathbf{k}}\hat{f}\hat{U}_{\mathbf{k}}$ (see Ref. 4). Now we introduce the heavy-fermion distribution matrix:

$$n^{ml}_{1\alpha\beta}(\mathbf{k},\mathbf{q},t)=\langle g^{\dagger}_{l\beta\mathbf{k}+\mathbf{q}/2}(t)g_{m\alpha\mathbf{k}-\mathbf{q}/2}(t)\rangle, \tag{7}$$

where the band indices m and l take values 1 and 2. Plugging Eq. (6) into (7), we arrive at a relationship between $\hat{N}_1$ and $\hat{n}_1$:

$$\hat{N}_1=\hat{U}^{-1}_{-}\hat{n}_1\hat{U}_{+}, \tag{8}$$

with $\hat{U}_{\pm}\equiv\hat{U}_{\mathbf{k}\pm\mathbf{q}/2}$. Here the diagonal components $n^{11}_{1\alpha\beta}(\mathbf{k},\mathbf{q},\omega)$ and $n^{22}_{1\alpha\beta}(\mathbf{k},\mathbf{q},\omega)$ describe the deviation of the quasiparticle distribution in the hybrid bands E_1 and E_2 from the equilibrium distribution. The off-diagonal components of $\hat{n}_1$ describe hybridization of quasiparticle states in the hybrid bands due to perturbations. Suppose that the total number of electrons per unit cell is smaller than two. Between the bands E_1 and E_2 there is a curved gap of order T_0, so that for $T\ll T_0$ only the lower band E_1 is partially filled, while the population of the upper band is negligible, i.e., $n^{22}_{1\alpha\beta}(\mathbf{k},\mathbf{q},\omega)\approx 0$.

Let us derive the transport equation for the heavy-fermion distribution function $n^{11}_{1\alpha\beta}(\mathbf{k},\mathbf{q},\omega)$. For convenience we introduce the following notation:

$$\delta n_{\alpha\beta\mathbf{k}}\equiv n^{11}_{1\alpha\beta}(\mathbf{k},\mathbf{q},\omega). \tag{9}$$

We start with the case of a zero external magnetic field, where the equilibrium distribution function is independent of the spin indices. Under deviations from equilibrium the spin density may fluctuate. Then the fluctuations $\delta n_{\alpha\mathbf{k}}\equiv\delta n_{\alpha\alpha\mathbf{k}}$ depend on the direction of spin, i.e., $\delta n_{\alpha}\neq\delta n_{\beta}$ for $\alpha\neq\beta$. Such a statement of the problem makes it possible to study spin diffusion and spin waves. Let $\varphi(\mathbf{r},t)$ be an external electric potential that generates an electric field $\mathbf{E}=-e\nabla_{\mathbf{r}}\varphi$ acting on the conduction electrons and the f-electrons. To allow for $\mathbf{E}$ in Eq. (1), we must add the term $e\varphi$ to the diagonal elements $\varepsilon^{cc}_{\alpha\alpha}$ and $\varepsilon^{ff}_{\alpha\alpha}$ of the quasiparticle energy matrix $\hat{\varepsilon}$. Linearizing the transport equation (1) in the matrix $\hat{N}_1$ and using Eq. (8), we obtain an equation for $\hat{n}_1$. Solving this equation with respect to the heavy-fermion distribution function $\delta n_{\alpha\mathbf{k}}$, we arrive at the desired transport equation:

$$\begin{aligned}&(\mathbf{q}\cdot\mathbf{v}-\omega)\,\delta n_{\alpha\mathbf{k}}-\mathbf{q}\cdot\mathbf{v}f'(E_{1\mathbf{k}})\\&\times\left(\frac{F^{s}_{0}}{2\rho^{*}_{F}}\sum_{\beta\mathbf{p}}\delta n_{\beta\mathbf{p}}+\frac{F^{a}_{0}\sigma_{\alpha}}{2\rho^{*}_{F}}\sum_{\beta\mathbf{p}}\sigma_{\beta}\delta n_{\beta\mathbf{p}}\right)\\&-ie\mathbf{E}\cdot\mathbf{v}f'(E_{1\mathbf{k}})=-iI(\delta n_{\alpha\mathbf{k}}),\end{aligned} \tag{10}$$

where $\sigma_{\alpha}=\pm 1$, $\mathbf{E}=\mathbf{E}(\omega,\mathbf{q})=-i\mathbf{q}\varphi(\omega,\mathbf{q})$, $\mathbf{v}=\nabla_{\mathbf{k}}E_{1\mathbf{k}}$ is the heavy-fermion velocity, $f(E)$ is the Fermi–Dirac distribution function, and $2\rho^{*}_{F}=2\rho_{F}m^{*}/m_0$ and m^{*} are the renormalized density of states with allowance for spin degeneracy and the heavy-fermion effective mass, respectively. Earlier a transport equation for $\delta n_{\alpha\mathbf{k}}$ without allowance of the interaction between heavy fermions was derived by Coleman[7] within a microscopic model of the Kondo lattice.

The Landau parameters F^{s}_{0} and F^{a}_{0} for $q\ll k_F$ and $\omega\ll T_0$ are given by the following simple relationships:

$$F^{s}_{0}=\frac{m^{*}}{m_0}-1\gg 1, \quad F^{a}_{0}=-\frac{T_m}{T_0},$$

where T_m characterizes the energy scale of the magnetic interaction between heavy fermions.[5] The large positive value of F^{s}_{0} means that there is strong potential repulsion between heavy fermions. Since heavy fermions are hybrid quasiparticles, there are two mechanisms for the magnetic interaction between heavy fermions. First, there is the indirect interaction via conduction electrons (the RKKY interaction). In this case T_m is positive and is of order the energy of the RKKY interaction of two spins localized at neighboring sites ($T_m=2N_fG^2\rho_0$, where G is the amplitude of the magnetic interaction between conduction electrons and f-electrons).[5] Here the Landau parameter F^{a}_{0} is negative, which corresponds to ferromagnetic spin–spin coupling between heavy fermions. Second, direct exchange interaction between localized f-electrons also contributes to the magnetic interaction between heavy fermions. If this interaction is antiferromagnetic and strong, the magnetic interaction between heavy fermions is antiferromagnetic ($F^{a}_{0}>0$).

To study galvanomagnetic phenomena, we write the transport equation for the case of a transverse magnetic field $\mathbf{H}\perp\mathbf{E}$. In order to allow for the Lorentz force acting only on the conduction electrons,

$$\mathbf{F}_L=\frac{e}{c}\mathbf{v}_0\times\mathbf{H},$$

where $\mathbf{v}_0=\nabla_{\mathbf{k}}\varepsilon(\mathbf{k})$ is the velocity of an electron in the conduction band with dispersion $\varepsilon(\mathbf{k})$, we must introduce the following substitution into Eq. (1):

$$\nabla_{\mathbf{r}}\varepsilon^{cc}_{\alpha\alpha}\to\nabla_{\mathbf{r}}\varepsilon^{cc}_{\alpha\alpha}-\mathbf{F}_L.$$

If we ignore spin, the transport equation is

$$(\mathbf{q}\cdot\mathbf{v}-\omega)\,\delta n_{\mathbf{k}}-\mathbf{q}\cdot\mathbf{v}f'(E_{1\mathbf{k}})\frac{F_0^s}{\rho_F^*}\sum_{\mathbf{p}}\delta n_{\mathbf{p}}-ie\mathbf{E}\cdot\mathbf{v}f'(E_{1\mathbf{k}})$$

$$-i\frac{e}{c}\mathbf{v}\times\mathbf{H}\cdot\nabla_{\mathbf{k}}\delta n_{\mathbf{k}}=-iI(\delta n_{\mathbf{k}}). \tag{11}$$

In the case of an isotropic Fermi surface for heavy fermions, this equation yields the well-known result for the Hall constant $R=1/ecN_t$, where $N_t=N_c+N_f$ is the total electron concentration. Generally, when studying galvanomagnetic phenomena, one must allow for changes in the topology of the Fermi surface resulting, first, from an increase in the number of electron states above the Fermi surface and, second, from the renormalization of the quasiparticle band. Indeed, for $T>T_0$ the Fermi surface is specified by the equation $\varepsilon(\mathbf{k})=\mu$, where the chemical potential μ is determined only by the conduction electron concentration N_c. When there is a transition to a heavy-fermion state ($T<T_0$), we have a quasiparticle band $E_{1\mathbf{k}}$, while the number of electrons above the Fermi surface is $N_t=N_c+N_f$. Obviously, in some cases there can be a transition from an electron Fermi surface to a hole Fermi surface, or an even more complicated transformation is possible.

3. SPIN WAVES

For normal metals the Landau parameters are assumed to be independent of frequency. For heavy-fermion metals such an approximation is valid only in the frequency range $\omega\ll T_0$. When $\omega>T_0$ holds, we must allow for the frequency dependence of the parameter F_0^a:

$$F_0^a(\omega)=\frac{F_0^aA(\omega)}{1-A(\omega)(|F_0^a|m_0/m^*)^{1/2}}, \tag{12}$$

where

$$A(\omega)=x^2(x^2-1)^{-1/2}\arctan(x^2-1)^{-1/2}, \tag{13}$$

and $x=2T_0\omega^{-1}\sqrt{m^*/m_0}$. At $\omega\propto 2T_0\sqrt{m^*/m_0}$ the function $F_0^a(\omega)$ changes sign. Such frequency behavior of $F_0^a(\omega)$ determines the spectrum of spin waves in the paramagnetic heavy-fermion state. The standard analysis of Eq. (10) (see, e.g., Ref. 6) leads to the following equation for the spin-wave spectrum:

$$\Gamma(q,\omega)=\frac{1}{F_0^a(\omega)}. \tag{14}$$

In the isotropic case $\Gamma(q,\omega)$ takes the form

$$\Gamma(q,\omega)=\frac{\lambda}{2}\ln\left|\frac{\lambda+1}{\lambda-1}\right|-1, \tag{15}$$

where $\lambda=\omega/qv_F$. If $F_0^a>0$ holds (antiferromagnetic interaction between heavy fermions), the spin-wave spectrum contains only an undamped mode ($\omega(q)=v_sq$) with a phase velocity v_s higher than the Fermi velocity of heavy fermions, v_F. If $F_0^a<0$ holds (ferromagnetic interaction between heavy fermions), Eq. (14) has a unique solution describing undamped spin waves

$$\omega(q)=\omega_0\left(1+\frac{4\pi}{3}(G\rho_0)^3\left(\frac{N_t}{N_f}\right)^2\left(\frac{q}{k_F}\right)^2\right)^{1/2} \tag{16}$$

with a finite excitation threshold

$$\omega_0=2T_0\sqrt{\frac{m^*}{m_0}} \tag{17}$$

and an extremely weak dependence on the wave vector q, since we assumed $G\rho_0\ll 1$. Equation (16) is valid at $T=0$.

4. SCATTERING OF HEAVY FERMIONS

Now let us discuss the scattering of heavy fermions and determine the collision integral $I(\delta n_{\alpha\mathbf{k}})$. In the case of pair collisions involving heavy fermions, we must calculate the amplitude $A_{\mathbf{k}\alpha,\mathbf{p}\beta}(\mathbf{q},\omega)$ for a process in which two heavy fermions with wave vectors $\mathbf{k}$ and $\mathbf{p}$ and spins α and β exchange momentum $\mathbf{q}$ and energy ω. To this end we must calculate (remaining within the scope of Eq. (10)) the polarization induced by a localized bare quasiparticle (see, e.g., Ref. 6). For $\omega\sim T$ and $T/v_F<q\ll k_F$ we obtain

$$A_{\uparrow\uparrow}=\frac{\pi^2}{k_Fm^*}\left(\frac{F_0^s}{1+F_0^s}+\frac{F_0^a}{1+F_0^a}\right), \tag{18}$$

$$A_{\uparrow\downarrow}=\frac{\pi^2}{k_Fm^*}\left(\frac{F_0^s}{1+F_0^s}-\frac{F_0^a}{1+F_0^a}\right).$$

The scattering probabilities for two heavy fermions with parallel spins ($W_{\uparrow\uparrow}$) and antiparallel spins ($W_{\uparrow\downarrow}$) are

$$W_{\uparrow\uparrow}=2\pi|A_{\uparrow\uparrow}|^2,\quad W_{\uparrow\downarrow}=2\pi|A_{\uparrow\downarrow}|^2.$$

Using the standard expression[6] for the reciprocal lifetime of a quasiparticle with an energy E such that $|E-\mu|\ll T$, we find that

$$\frac{1}{\tau(E)}=\frac{(m^*)^3}{16\pi^4}\left\langle\frac{W}{\cos(\theta/2)}\right\rangle(\pi^2T^2+(E-\mu)^2)$$

$$=\frac{\pi y}{32}\,\frac{\pi^2T^2+(E-\mu)^2}{\varepsilon_F^*}, \tag{19}$$

where the angle brackets stand for averaging over angles. Moreover, we have $\varepsilon_F^*\equiv k_F^2/2m^*$ and

$$y=\left(\frac{F_0^s}{1+F_0^s}+\frac{F_0^a}{1+F_0^a}\right)^2+2\left(\frac{F_0^s}{1+F_0^s}-\frac{F_0^a}{1+F_0^a}\right)^2. \qquad (20)$$

According to Eq. (10), the collision of two heavy fermions depends on both the potential interaction and the spin–spin coupling. If we ignore the latter, i.e., put $F_0^a=0$, and allow for the fact that $F_0^s\gg 1$, we get $y=3$. In this case, Eq. (19) coincides (including the numerical factor) with the expression for the rate of scattering of heavy fermions by slave bosons found in the microscopic studies of Kondo lattices.[7–10] This means that the microscopic approach used in Refs. 7–10 to study the scattering of heavy fermions by slave bosons is simply a convenient mathematical representation of the collision of heavy fermions. When there is strong competition between magnetism and the Kondo effect, in other words, when the system is close to the threshold of ferromagnetic instability of the heavy-fermion state ($T_0\sim T_m$), i.e., at $F_0^a\sim -1$, we have $y\gg 3$, and spin–spin coupling provides the main contribution to $\tau(E)$.

For normal metals, Eq. (19) contains the Fermi energy $E_F\equiv k_F^2/2m_0=\varepsilon_F^* m^*/m_0$ instead of $\varepsilon_F^*\equiv k_F^2/2m^*$. This implies that the pair collision rate for heavy fermions is much larger than the pair collision rate for electrons in normal metals, i.e.,

$$\frac{1}{\tau(E)}=\frac{1}{\tau_0(E)}\frac{m^*}{m_0}\gg\frac{1}{\tau_0(E)}.$$

The pair collision rate determines the coefficient A in the temperature dependence of resistance ($R=R_0+AT^2$). This is precisely why for heavy-fermion metals this coefficient is anomalously large (the expression for A can be found in Refs. 7–10).

Now let us examine the collision of heavy fermions with charged impurities. According to the theory of normal Fermi liquids,[6] the effective potential energy of a quasiparticle in the field of a point impurity with charge Ze and with allowance for screening effects is

$$V_{\mathbf{q}}^{\text{eff}}=\frac{4\pi Ze^2}{q^2}\frac{1}{\varepsilon(\mathbf{q},0)}\frac{1}{1+F_0^s}. \qquad (21)$$

This quantity is also equal to the matrix element for the quasiparticle transition from a state with a wave vector $\mathbf{k}$ to a state with a wave vector $\mathbf{k}+\mathbf{q}$. The effective potential energy $V_{\mathbf{q}}^{\text{eff}}$ for $\mathbf{q}$ small can be found by using the static dielectric function:[5,11]

$$\varepsilon(\mathbf{q},0)=\varepsilon_0+\frac{4\pi e^2}{q^2}\frac{2\rho_F^*}{1+F_0^s}. \qquad (22)$$

This leads to the simple formula

$$V_{\mathbf{q}}^{\text{eff}}=\frac{Z}{2\rho_F^*}, \qquad (23)$$

which shows that the effective impurity potential in heavy-fermion systems is lower by a factor of m_0/m^* than the same potential in normal metals. Using the standard electron–impurity collision integral,[9] we find that for impurity scattering the lifetime τ_i of heavy fermions increases by a factor of m^*/m_0 in comparison to the lifetime in normal metals: $\tau_i=\tau_{0,i}m^*/m_0\gg\tau_{0,i}$. This result follows from the fact that $\tau_i^{-1}\propto\rho_F^*|V^{\text{eff}}|^2$. In other words, for a given charged-impurity concentration, the rate at which heavy fermions collide with impurity ions is m^*/m_0 times lower than the rate at which electrons collide with impurities in normal metals. This is in full agreement with the results of microscopic theory (see, e.g., Ref. 9). The impurity scattering rate determines the value of the residual resistance, which is of the same order as in normal metals. Indeed,

$$\sigma\propto e^2\rho_F^* v_F^2\tau_i=e^2\rho_0 v_{0F}^2\tau_{0,i}\propto\sigma_0.$$

5. DISCUSSION AND CONCLUSION

We have used the Landau theory of normal Fermi liquids to derive a transport equation for heavy fermions and have found that the equation is of standard form. Hence to describe transport phenomena in heavy-fermion compounds, one can employ the methods developed for the kinetics of normal Fermi liquids. What is important is that the derived transport equation takes into account the potential and magnetic interactions between heavy fermions. We have found that the amplitudes of these interactions (or, in other terms, the symmetric and antisymmetric Landau parameters) depend on frequency. However, in the limit $\omega\ll T_0$ this frequency dependence can be ignored. It can play an important role only when $\omega>T_0$. In particular, this determines the spectrum of spin waves in a paramagnetic heavy-fermion state.

We have established that the spin-wave spectra for compounds with ferromagnetic and antiferromagnetic interactions between heavy fermions differ considerably. In the case of antiferromagnetic interaction there are only gapless acoustic-like spin waves, $\omega(q)=v_s q$, with a phase velocity higher than the Fermi velocity of the heavy fermions. But if the interaction is ferromagnetic, the Fermi-liquid approach predicts the presence of a gap in the spin-wave excitation spectrum, a gap whose width is proportional to the low-temperature Kondo scale. Moreover, these spin waves have exceptionally weak dispersion. Spin waves with a gap exist only in a heavy-fermion state, i.e., at temperatures $T<T_0$. As is known, the magnetic susceptibility of a heavy-fermion metal has a peak in the temperature range of a transition from a state with incoherent Kondo scattering to a coherent state with heavy fermions. Hence in experiments one should expect gap spin waves to disappear at temperatures above the value at which the magnetic susceptibility is at its maximum. Recent inelastic neutron scattering studies in the $YbAl_3$ (Ref. 12) and CeNi (Ref. 13) compounds, which belong to the class of valence-fluctuation compounds, discovered a gap in the magnetic response that exists only at temperatures below the temperature of the maximum in the magnetic susceptibility, i.e., in a situation that qualitatively resembles the one discussed earlier. Another important similarity is the proximity of the energy of the observed gap to the Kondo temperature for these compounds. Unfortunately, at present there is no irrevocable evidence in favor of the formation of hybrid

quasiparticle bands in these compounds. This makes it impossible to decide whether the gap in the spectrum of spin excitations predicted by the Landau theory is the gap discovered by Murani[12] and Lazukov *et al.*[13]

We have established that in a heavy-fermion state the galvanomagnetic properties are also determined from a standard transport equation. What is important here (and this fact should be stressed in connection with studies of properties of heavy-fermion metals) is the renormalization of the Fermi surface and, possibly, a change in the surface's topology as a result of a transition to the heavy-fermion state.

Remaining within the scope of the standard Fermi-liquid approach, we have found the rate of pair collision of heavy fermions and the rate of scattering of these quasiparticles by impurities. The rate of pair collisions of heavy fermions coincides with the rate of scattering of heavy fermions by slave bosons found in the microscopic of heavy-fermion compounds. This means that the scattering of heavy fermions by slave bosons is only a convenient mathematical tool for describing collisions of heavy fermions due to the potential and magnetic interactions between the heavy fermions. The results of this study of the impurity scattering of heavy fermions also fully agree with those of the microscopic theory.

The present work was supported by the Russian Fund for Fundamental Research (Grant No. 96-02-16848) and the program ISI's Network-Esprit (I.S.I. Foundation, Torino).

*)E-mail: goltsev@gav.ioffe.rssi.ru

[1] N. B. Brand and V. V. Moshchalkov, Adv. Phys. **33**, 373 (1984).
[2] N. Grewe and F. Steglich, in *Handbook of Physics and Chemistry of Rare Earths*, Vol. 14, K. A. Gschneider, Jr., and LeRoy Eyring, North-Holland, Amsterdam (1991), p. 343.
[3] D. M. Newns and N. Read, Adv. Phys. **36**, 799 (1987).
[4] A. V. Gol'tsev and V. V. Krasil'nikov, JETP Lett. **61**, 287 (1995); J. Phys.: Condens. Matter **7**, 6523 (1995).
[5] A. V. Gol'tsev, JETP Lett. **62**, 223 (1995); J. Phys.: Condens. Matter **8**, 457 (1996).
[6] D. Pines and F. Nozieres, *Theory of Quantum Liquids*, W. A. Benjamin, Reading, MA (1966).
[7] P. Coleman, Phys. Rev. Lett. **59**, 1026 (1987).
[8] A. Auerbach and K. Levin, Phys. Rev. Lett. **57**, 877 (1986).
[9] A. J. Millis and P. A. Lee, Phys. Rev. B **35**, 3394 (1987).
[10] V. I. Belitskiĭ and A. V. Gol'tsev, Zh. Éksp. Teor. Fiz. **96**, 1815 (1989) [Sov. Phys. JETP **69**, 1026 (1989)]; Physica B **172**, 459 (1991).
[11] A. J. Millis, M. Lavagna, and P. A. Lee, Phys. Rev. B **36**, 864 (1987).
[12] A. P. Murani, Phys. Rev. B **50**, 9882 (1994).
[13] V. N. Lazukov, P. A. Alekseev, E. S. Clementyev, R. Osborn, B. Rainford, I. P. Sadikov, O. D. Chistyakov, and N. B. Kolchugina, Europhys. Lett. **33**, 141 (1996).

Translated by Eugene Yankovsky

Anisotropy of the transport properties of single-crystal Bi_2Te_3 disordered by electron bombardment

A. E. Kar'kin,*) V. V. Shchennikov, B. N. Goshchitskiĭ, S. E. Danilov, and V. L. Arbuzov

Institute of Metal Physics, Ural Branch of the Russian Academy of Sciences, 620219 Ekaterinburg, Russia

(Submitted 30 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1787–1798 (May 1998)

Using single-crystal samples of Bi_2Te_3 bombarded by 5-MeV electrons at a temperature of 250 K, we study the electrical resistivity and the Hall effect in the temperature range 1.7–370 K and the Shubnikov–de Haas effect at $T=4.2$ K in magnetic fields up to 14 T. We find that electron bombardment of Bi_2Te_3 crystals results in a transition from the metallic p-type state to the metallic state with a Fermi surface. Annealing at 350 K eliminates the radiation defects and restores the p-type metallic conductance. © *1998 American Institute of Physics.* [S1063-7761(98)01605-9]

1. INTRODUCTION

High-T_c superconductors in a narrow concentration range near the metal–insulator transition[1,2] exhibit an anomalous dependence, which disagrees with the Fermi-liquid description,[3] of the resistivity ρ and the Hall constant R on temperature,[4] level of doping,[5] and degree of disorder.[2,6] For instance, the behavior of ρ and R in the ab plane and that along the c axis differ so dramatically that the anisotropy of these effects becomes strongly temperature-dependent. This suggests that theoretical and experimental studies of high-T_c superconductors near the metal–insulator transition are very important.[3,7] To be able to experimentally investigate the properties of high-T_c superconducting materials in the normal state at low temperatures, superconductivity must be suppressed, which is possible only in a very strong magnetic field, $B=30$–100 T. It is therefore preferable to study low-dimensional nonsuperconducting model systems (or systems with low transition temperatures) as functions of the doping level or degree of disorder. Studies of this kind involving the anisotropic KWO_3 and $NbSe_2$ compounds disordered by low-temperature neutron bombardment made it possible to detect correlations between the variations of the electronic properties and the degree of disorder in a metal–insulator transition.[8]

In this paper we use single crystals of Bi_2Te_3 as a model system. The crystals are bombarded by fast electrons and are then annealed in order to change the charge carrier concentration. The Bi_2Te_3 compound has been thoroughly studied both theoretically[9,10] and experimentally[11,12] as a promising material for thermoelements.[13–15] The reason this material was chosen for our studies is that it is a narrow-gap heavily doped semiconductor in which the anisotropy of electrical properties resembles that of high-T_c superconducting materials.[2,8] The high perfection of the single-crystal samples makes it possible to detect quantum magnetic oscillations and use these oscillations to find the parameters of the electron spectrum.[11,12]

We study the anisotropy of the electrical resistivity and of the Hall effect, as well as the Shubnikov–de Haas effect. The behavior of the anisotropy parameter of galvanomagnetic effects (a parameter that characterizes the motion of charge carriers in different directions) in low-dimensional systems correlates with the changes in the electron spectrum near a metal–insulator transition.[2,8] The aim of our paper is to investigate the features of the electronic states that arise in the Bi_2Te_3 system when disorder is introduced into the crystalline structure by electron bombardment.

2. SAMPLES AND THE MEASUREMENT METHOD

The galvanomagnetic properties of single-crystal samples of Bi_2Te_3 were studied in the temperature interval 1.7–370 K in a constant magnetic field up to 14 T using an Oxford Instruments device. The electrical contacts manufactured by ultrasonic indium soldering measured 50 by 50 μm. The electrical resistivities ρ_a and ρ_c and the Shubnikov–de Haas effect were measured by Montgomery's method[16] using samples that measured 1 by 0.5 by 0.2 mm. The Hall constants R_{ac} and R_{ca} (the first subscript corresponds to the direction of the current and the second, to the direction of the magnetic field) were measured using, respectively, 1-by-1-0.05 mm^3 and 1-by-0.2-by 0.2 mm^3 samples with four symmetric contacts. In the process the directions of the measuring current and the magnetic field B were changed and the current and potential currents were switched (which is equivalent to changing the direction of the field B). In contrast to the standard methods, in Montgomery's method[16] ρ_a and ρ_c are measured in a single sample, which reduces the possible errors in estimating the anisotropy of the electrical resistivity.[17]

The bismuth telluride samples were bombarded with 5-MeV electrons in the chamber of an electron accelerator, which was accompanied by cooling with gaseous helium to a temperature below 250 K. The measurements were done using the initial samples (denoted B1), after irradiation with a fluence $\Phi=1\times10^{19}\text{cm}^{-2}$ (B2), and after isochronous annealing at 350 K (B3) and 375 K (B4) in the course of

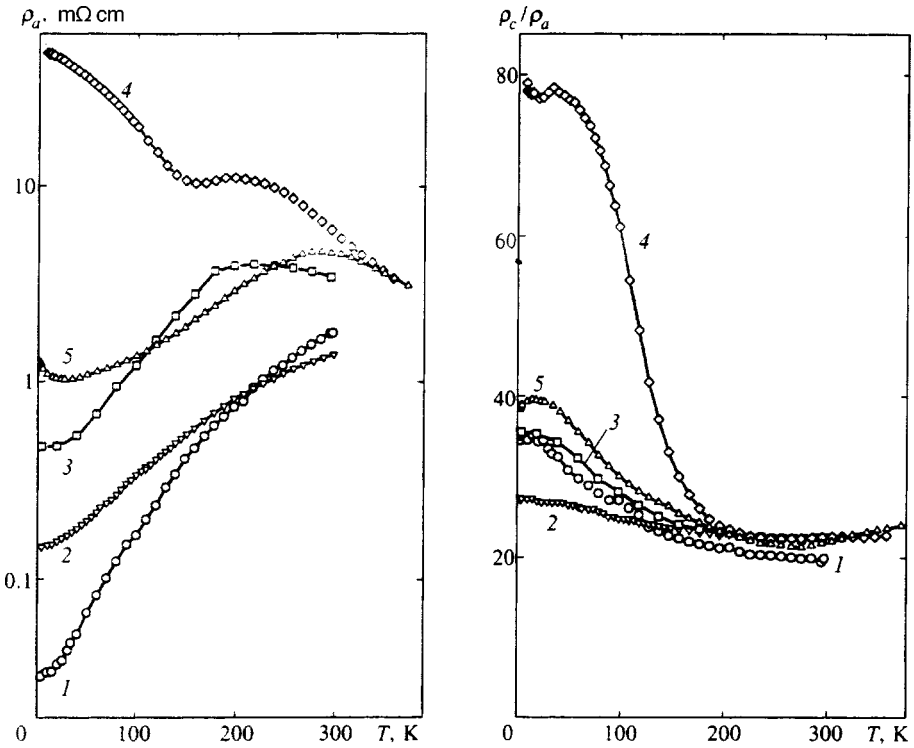

FIG. 1. Temperature curves for ρ_a and the resistivity anisotropy ρ_c/ρ_a of the bismuth telluride samples: curve *1*, the initial sample B1; curve *2*, the sample B2 bombarded by 5-MeV electrons at $T=250$ K with a fluence $1\times10^{19}\text{cm}^{-2}$; curves *3* and *4*, the samples bombarded by electrons and annealed at $T=350$ K and 375 K (B3 and B4) for 20 min; and curve *5*, the sample annealed at 375 K (B5) for 120 min.

20 min and at 375 K (B5) in the course of 120 min. No annealing was done at higher temperatures because of the possibility of indium diffusion into the sample.[15]

3. RESULTS OF MEASUREMENTS

Figure 1 depicts the temperature curves for the electrical resistivity ρ_a and the resistivity anisotropy ρ_c/ρ_a, while Fig. 2 depicts the temperature curves for the Hall constants R_{ac} and R_{ca} of the irradiated crystals. At first glance the effect of electron bombardment on ρ and R is equivalent to introducing indium into the lattice:[11] the hole concentration decreases, the resistance grows, and the ρ (T) dependence resembles that of a semiconductor at $T>250$ K. Annealing (sample B4) lowers the electron concentration, so that the semiconductor behavior of $\rho(T)$ can be observed at all temperatures T with the exception of the liquid-helium range,

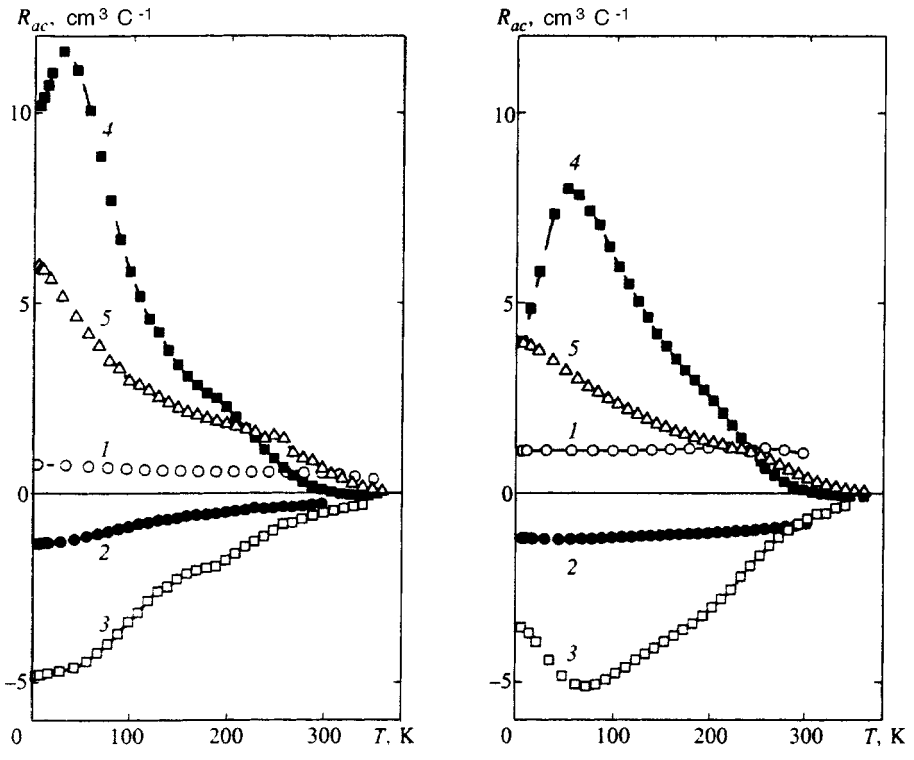

FIG. 2. Temperature curves for the Hall constants R_{ac} ($\mathbf{j}\|\mathbf{a}$, $\mathbf{B}\|\mathbf{c}$) and R_{ca} in a magnetic field $B=12$ T for samples B1–B5. The notation for the curves is the same as in Fig. 1.

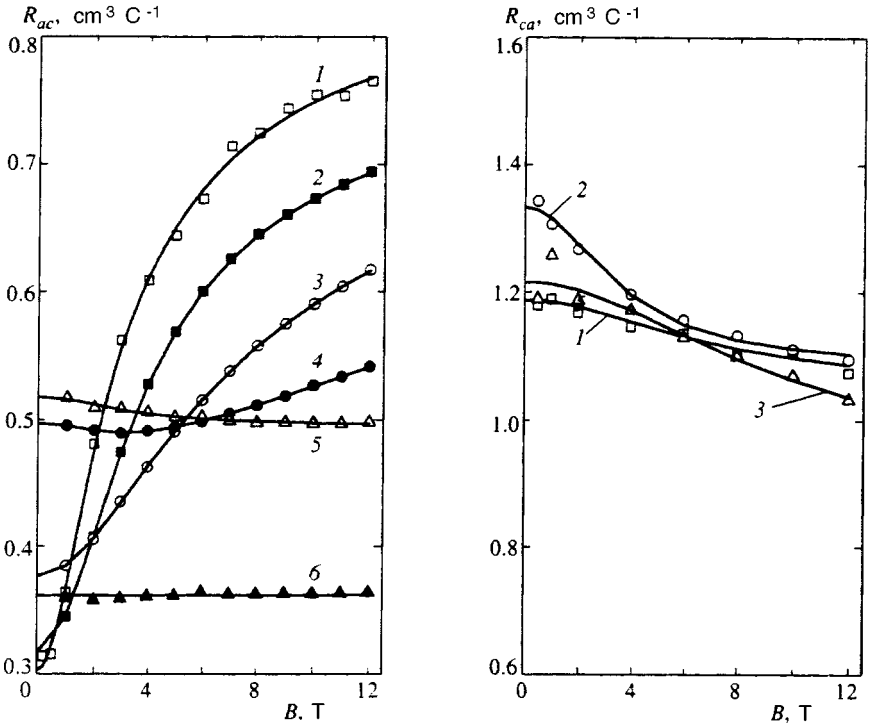

FIG. 3. Hall constants R_{ac} and R_{ca} as functions of the magnetic field B for the initial Bi_2Te_3 sample (B1) at different temperatures: curve *1* — 4.2 K; curve *2* — 50 K; curve *3* — 100 K; curve *4* — 188 K; curve *5* — 300 K; and curve *6* — 350 K. □, ■, ○, ●, △, and ▲ represent the experimental data, and the curves are fitting curves (see the text).

and there is a minimum in ρ at $T \approx 150$ K. The emergence of a section of activation dependence at $T < 150$ K corresponds to a sharp increase in anisotropy in this temperature interval. At 4.2 K the resistivity of the sample B5 is 40 times higher than that of the initial sample B1, i.e., under such annealing the properties are not completely restored.

Electron bombardment of Bi_2Te_3 samples changes the sign of the Hall constants R_{ac} and R_{ca} (Fig. 2, sample B2). After annealing (samples B3–B5) there is a partial recurrence of the galvanometric properties to the initial state. The variations in the carrier concentrations with temperature depicted in Fig. 2 can be described qualitatively by the following expressions for an extrinsic semiconductor:[18]

$$n = \frac{n_0}{2} + \left[\left(\frac{n_0}{2} \right)^2 + (4A_cA_v)^2 \exp\left(-\frac{2E_g}{kT} \right) \right]^{1/2},$$
$$p = -\frac{n_0}{2} + \left[\left(\frac{n_0}{2} \right)^2 + (4A_cA_v)^2 \exp\left(-\frac{2E_g}{kT} \right) \right]^{1/2}, \quad (1)$$

where n_0 is the difference in the concentrations of the donor and acceptor impurities, A_c and A_v are the densities of states in the conduction and valence bands, and E_g is the band gap. Electron bombardment changes the value and sign of n_0, which is equivalent to doping. Annealing reduces the nonequilibrium value of n_0 reached as a result of bombardment, with the curves in Fig. 2 reflecting this fact. Equations (1) provide a good description of the behavior of R for the samples B1 and B2 over the entire temperature range and for the samples B3–B5 at high temperatures. At low temperatures, as Figs. 1 and 2 show, the samples B3–B5 acquire a new activation dependence with lower activation energy. The anisotropy of R for B1 and B2 corresponds to the topology of the electron structure,[15] for which[13–15]

$$n \simeq \frac{0.3}{R_{ac}e} \simeq \frac{0.6}{R_{ca}e}. \quad (2)$$

This relationship does not hold for the samples B3–B5.

The characteristic scale of variations of R with temperature increases from sample B2 to sample B5; an estimate sets the increase in the activation energy E_g at 0.05 to 0.15 eV, respectively. For B4 and B5 at room temperatures, the Hall constants change sign; such behavior, due to the different temperature dependence of the concentrations and mobilities of electrons and holes, was observed in *n*-type samples with excess tellurium content.[19]

The Hall constants R_{ac} have a strong field dependence (Figs. 3 and 4), which points (just as the temperature dependence of the Hall constants does) to the fact that there are several groups of carriers with different concentrations n_i and mobilities μ_i participating in conduction. Analogous behavior of the $R(B)$ dependence, related to the participation of thermally activated electrons in the process of charge transfer, was observed by Azou *et al.*[11] The similar dependence for R_{ca} is extremely weak at all temperatures, which suggests that the contribution of holes of only one type is predominant. Even in the case of two-band conductance, exact analysis of the $R(B)$ and $\rho(B)$ behavior is difficult: in Bi_2Te_3 one must allow for three groups of carriers[13–15] and, moreover, the material is anisotropic. To fit the experimental curves, we used the following expressions for an isotropic semiconductor:[20]

$$R = \frac{\langle R\sigma^2 \rangle}{\langle \sigma \rangle^2 + (\langle R\sigma^2 \rangle B)^2}, \quad (3)$$

$$\rho = \frac{\langle \sigma \rangle}{\langle \sigma \rangle^2 + (\langle R\sigma^2 \rangle B)^2}, \quad (4)$$

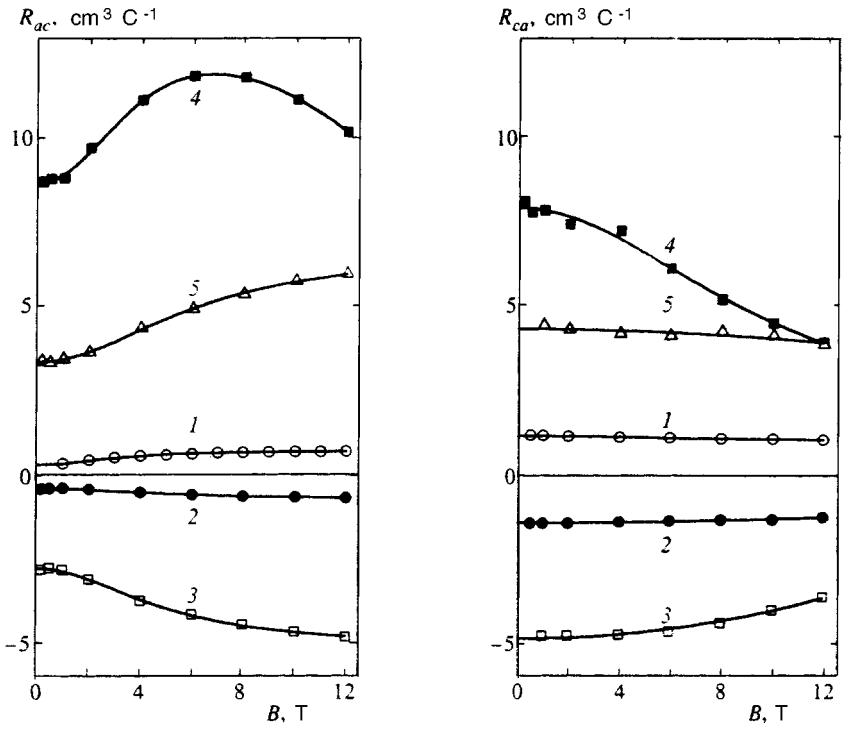

FIG. 4. Hall constants R_{ac} and R_{ca} as functions of the magnetic field B at $T=4.2$ K for samples B1–B5. The notation of the curves is the same as in Fig. 1. ■, △, ○, ●, and □ represent the experimental data, and the curves are fitting curves (see the text).

where

$$\langle\sigma\rangle=\sum_i \frac{\sigma_i}{1+(R_i\sigma_i B)^2}, \quad \langle R\sigma^2\rangle=\sum_i \frac{R_i\sigma_i^2}{1+(R_i\sigma_i B)^2}, \tag{5}$$

with $\sigma=1/\rho_i$ and R_i the conductivities and Hall constants for individual bands. For two-band conduction, Eqs. (3)–(5) lead to the well-known expressions discussed in Ref. 21. We ignored the field dependence of the one-band parameters σ_i and R_i, since we assumed the bands to be strongly degenerate (the Hall factors are equal to unity).

The results of fitting the experimental data on the $R_{ac}(B)$ and $\rho_a(B)$ dependence measured at different temperatures (Fig. 3) for the sample B1 to the values obtained via Eqs. (3)–(5) under the assumption that there are two or three (when necessary) groups of carriers are listed in Table I. In the first stage we fitted the experimental data on $R_{ac}(B)$ with a fixed value of conductivity at $B=0$. The obtained parameters were then used as initial ones to fit the experiment data on the $\rho_a(B)$ dependence. The parameters obtained as a result of these two procedures were found to differ by 10–30%. They were refined to achieve the best possible description of both experimental dependences. The carrier parameters σ_i and R_i for two orientations (along the **c** axis and at right angles to **c**) were chosen independently.

Processing the experimental data via Eqs. (3)–(5) showed that in the initial Bi_2Te_3 sample the charge transfer in the plane is done by electrons and holes whose mobility values are close at all temperatures but that the hole contribution to conductivity is predominant. The partial conductivity of the third group of carriers is low, and at low temperatures the mobility of these carriers, $\mu_3=\sigma_3 R_3$, is much lower than the mobility of the first two groups and is weakly temperature-dependent. Along the **c** axis the charge is transferred only by holes, and the electrons have an extremely low mobility. The fitting results agree with the temperature behavior of concentrations for the three-band model suggested in Ref. 19. Analogous behavior of R_{ac} and R_{ca} and temperature and field variations was observed by Testardi *et al.*[22]

A similar fitting of the field dependence of R_{ac}, R_{ca} (see Fig. 4), ρ_a, and ρ_c made it possible to obtain a set of parameters σ_i and R_i for the irradiated crystals (Table II). For sample B2 the main contribution to the conductivity and the

TABLE I. One-band parameters σ_i and R_i $(i=1,2,3)$ obtained in fitting the experimental curves for the Hall constants $R(B)$ and the electrical resistivities $\rho(B)$ in Eqs. (3)–(5) for the initial Bi_2Te_3 sample B1 at different temperatures T and field orientations with respect to the crystallographic axes a and c.

	B∥c						**B∥a**	
	σ_i, $10^6(\Omega\text{m})^{-1}$			R_i, $10^{-6}\text{m}^3\text{C}^{-1}$				
T,K	σ_1	σ_2	σ_3	R_1	R_2	R_3	σ_1, $10^6(\Omega\text{m})^{-1}$	R_1, $10^{-6}\text{m}^3\text{C}^{-1}$
4.2	2.5	0.5	0.03	+0.5	−1.6	−6	0.06	1.2
50	1.2	0.2	0.01	+0.5	−2	−6	0.05	1.1
100	0.5	0.07	–	+0.6	−4	–	0.02	1.1
300	0.05	0.007	–	+0.7	−4	–	0.003	0.9

TABLE II. One-band parameters σ_i and R_i obtained in fitting the experimental curves for $R(B)$ and $\rho(B)$ in Eqs. (3)–(5) for the Bi_2Te_3 sample B2–B5 at different orientations of field B.

	B$\|$**c**						**B**$\|$**a**			
	σ_i, $10^6(\Omega\text{m})^{-1}$			R_i, $10^{-6}\text{m}^3\text{C}^{-1}$			σ_i, $10^6(\Omega\text{m})^{-1}$		R_i, $10^{-6}\text{m}^3\text{C}^{-1}$	
Samples	σ_1	σ_2	σ_3	R_1	R_2	R_3	σ_1	σ_2	R_1	R_2
B2	0.1	0.6	–	+3	−0.6	–	0.005	0.02	−20	−1
B3	0.02	0.2	–	+15	−4	–	0.003	0.003	+30	−40
B4	0.001	1×10^{-4}	0.001	+80	−20	−2000	8×10^{-6}	8×10^{-6}	$+7\times10^3$	-6×10^3
B5	0.08	0.01	–	+5	−20	–	0.001	0.001	+30	−10

Hall effect is provided by electrons, while the hole contribution is either small (the measuring current **j** is directed along the **a** axis) or nil (**j**$\|$**c**). In contrast to R_{ac}, the value of R_{ca} is essentially independent of temperature, which reflects the difference in carrier type for the directions along **a** (electrons and holes) and along **c** (only electrons). For all the samples, $R_{ca}(B)$ is described by a model with two types of carrier. The nonmonotonic dependence of R_{ac} on B (sample B4), which has a maximum for $B>6$ T, required three types of carrier for fitting: two with relatively high mobilities and one with low mobility (Table II). The sample B4 at $T=4.2$ K has high values $\rho_a\approx 50\ \text{m}\Omega\cdot\text{cm}$ and $\rho_c/\rho_a\approx 80$ (Fig. 1); high-mobility electrons and holes provide the main contribution to conductivity for the case where **j**$\|$**a** and $B=0$, holes with a lower mobility provide the main contribution for the case where **j**$\|$**c** (Table II).

At $T=4.2$ K, oscillations in ρ_a in transverse and longitudinal magnetic fields up to 14 T were observed in the initial sample B1. Of the irradiated samples only B5 (the sample with the smallest disorder) exhibited oscillations in ρ_a in a transverse field (Fig. 5). To fit the experimental curves we used the expression[20]

$$\rho_a(B)=\rho_0(B)\left[1+a_0\ \exp\left(-\frac{B_0}{B}\right)\sin\left(\frac{B_1}{B}+\varphi\right)\right]^{-1}, \quad (6)$$

where the monotonic part $\rho_0(B)$ (classical magnetoresistance) is described by Eqs. (4) and (5), the factor $a_0\exp(-B_0/B)$ takes into account the damping of oscillations (B_0 is related to mobility by the formula $\mu_0=\pi/B_0$), and B_1 is determined by the extremal cross-sectional area of the Fermi surface: $S_F=2\pi eB_1/\hbar$. Processing the experimental curves of longitudinal magnetoresistance required allowing for the contribution of two oscillations with distinct periods. The parameters obtained in the fitting process are listed in Table III. The tendency of μ_0 and B_1 in B5 to be smaller than the values in the initial sample B1 on the whole agrees with the behavior of the mobilities $\mu_1=\sigma_1R_1$ and concentrations $n_1\sim 1/R_1e$ (see Tables I–III), but μ_0 is smaller than μ_1. The quantity m_0, which determines the damping of oscillations, is related to the cyclotron mass and in the anisotropic case does not coincide with the Hall mobility μ_1. If one uses the parameters of Azou *et al.*,[11] the result is $\mu_1/\mu_0\approx 2$. For B1 and B5 this ratio is approximately four,

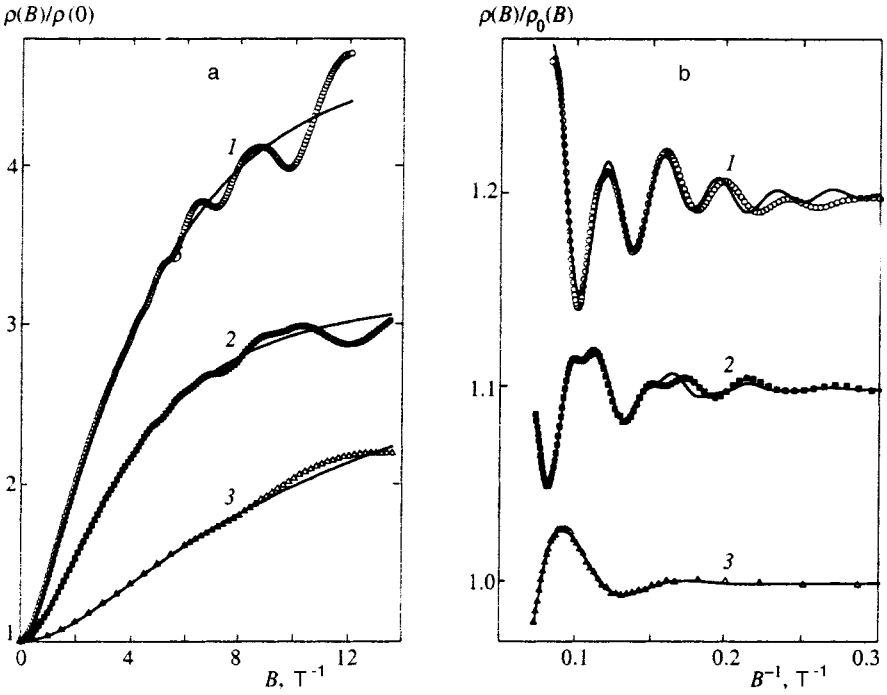

FIG. 5. Oscillations of the transverse (curves *1* and *3*) and longitudinal (curve *2*) magnetoresistances $\rho(B)$ in Bi_2Te_3; ρ_0 is the monotonic part of the resistivity (see Eq. (6)); curves *1* and *2*, the initial sample B1; and curve *3*, the irradiated sample B5. ○, ■, and △ represent the experimental data, and the curves are fitting curves for $\rho_0(B)$ (a) and the oscillating part of the resistivity (b, the curves are shifted along the horizontal axis).

TABLE III. Parameters obtained in fitting the oscillations of the magnetoresistance $\rho_a(B)$ (the Shubnikov–de Haas effect to Eq. (6) for the initial (B1) and irradiated (B5) Bi_2Te_3 samples.

Sample	Field orientation	B_0,T	$\mu_0=\pi/B_0$	B_1,T	B_2,T
B1	**B**$\|$**c**	15	0.2	170	–
B2	**B**$\|$**a**	15	0.2	260	120
B5	**B**$\|$**c**	35	0.09	80	–

which is probably due to the higher anisotropy of the electrical resistivities of our samples as compared to the *n*-type samples used by Azou *et al.*[11]

The periods of oscillations B_x, B_y, and B_z for the three crystallographic directions can be used to calculate the carrier concentration:

$$n=6\frac{4}{3\pi\sqrt{2}}\left(\frac{e}{h}\right)^{3/2}(B_xB_yB_z)^{1/2}$$

(for six ellipsoids). For B1 we have $n\approx 4\times 10^{18}\text{cm}^{-3}$, while from the Hall concentration (Eq. (2)) we obtain $n\approx 5\times 10^{18}\text{cm}^{-3}$. Estimates for B5 yield $n\approx 1.5\times 10^{18}\text{cm}^{-3}$ and $0.8\times 10^{18}\text{cm}^{-3}$, respectively.

4. DISCUSSION

Ordinarily, the Bi_2Te_3 compound with a C-33 layered structure has an excess of positive charge carriers, which form as a result of small deviations from the stoichiometric composition in the sample crystallization process.[9,12] The rhombohedral crystalline structure of bismuth telluride consists of layers of tellurium and bismuth that alternate in the sequence Bi–Te–Te–Bi–Te–Bi–Te–Te–Bi(Refs. 13 and 14), with the unit cell containing one Bi_2Te_3 molecule. According to Ref. 9, the hole and electron constant-energy surfaces consist of six ellipsoids oriented at, respectively, 86°- and 56°-angles to the *c* axis. This band structure, which can be considered a first approximation, determines the anisotropy of the resistivities ρ_a and ρ_c and the Hall constants R_{ac} and R_{ca}. Some experimental data suggest, however, that the top of the valence band and the conduction band have a more complicated structure.[15] The three-band model proposed in Ref. 19 provides a good description of the temperature anomalies in *R*, including sign inversion. Calculations of the antistructural defect in Bi_2Te_3 by the Green's functions method[23] supported the version of the valence band used in the model of Ref. 19. According to Pechur and Toussaint,[23] the top of the valence band has maxima that differ by 0.02 eV in the directions ZG, ΓA, ΓD, and ΓX of the Brillouin zone. The minimum in the conduction band is observed along the ΓZ direction, where the energy changes little. Preliminary analysis supports this picture.

When Bi_2Te_3 is bombarded by 5-MeV electrons, defects form in the compound, effectively doping the material. Our estimates suggest that the relative number of initially knocked-out atoms or dislocations (point defects of the type of vacancies and interstitial atoms) amounts to 10^{-3}–10^{-4} for an electron fluence $\Phi=1\times 10^{19}\text{cm}^{-2}$. Since the irradiation temperature of 250 K is relatively high (approximately three-tenths of the melting point, equal to 860 K; Ref. 24), most initial point defects are annealed in the irradiation process. The above data suggest that the main type of the radiation-induced defects in this case is the ''antistructural'' defects, in which Bi and Te atoms change places; these defects are present in the initial samples.[15] More complicated defects (complexes) may also be preserved at $T=250$ K. The concentration of antistructural defects is of order 10^{-4}, which must lead to a corresponding increase in electron and hole concentrations by approximately 10^{18}cm^{-3}.

Antistructural defects are usually assumed to be responsible for the hole concentration prevailing in Bi_2Te_3 (Ref. 24). The thermodynamically stable concentration of such defects yields 10^{19} holes per cubic centimeter in Bi_2Te_3 (Ref. 19). The excess of defects introduced by irradiation lowers the hole concentration and increases the electron contribution to conductivity, while annealing reduces the defect concentration to the thermodynamically stable value at the given temperature and hence increases the hole contribution. The change of sign of the thermoelectromotive force and the Hall effect when *p*-Bi_2Te_3 is bombarded by neutrons and protons has been observed earlier (see the literature cited in Ref. 15). Note that applying pressure up to 20 GPa, which initiates structural phase transitions in Bi_2Te_3 and in an indium alloy,[25] also leads to a transition from *p*-type conduction to *n*-type, i.e., in this case the introduced defects lower the hole concentration.

In the initial (B1) and irradiated (B2) samples, the dominant charge carriers are mobile electrons and holes. The only difference between B1 and B2 is that in the former the Fermi level is close to the lower hole band, while in the latter the Fermi level is near the upper electron band. This band structure for Bi_2Te_3 was used in Ref. 19. The decrease in the concentration of radiation-induced defects under isochronous annealing ensures a transition from *n*-type conduction to *p*-type, so that the Fermi level lands in the region near the middle of the gap (sample B4). An important feature of this state is that the concentrations of mobile electrons and holes are extremely low (Table II), so that the contribution of low-mobility carriers becomes significant.

Since in the B1 sample the main contribution to conductivity for all values of *T* is provided by holes and in the B2 sample, by electrons (see Tables I and II), we can compare the mobilities by defining them as $\mu\simeq R_{ac}/\rho_a$. The temperature dependence of the reciprocal mobilities μ^{-1} is depicted in Fig. 6 and is described fairly well by the formula

$$\mu^{-1}=a+bT^n, \tag{7}$$

where $n=2.2\pm 0.2$. The term *a*, which corresponds to ''impurity'' scattering, increases under bombardment, while the temperature-dependent term change very little. The exponent *n* is close to two, which is characteristic of electron–electron scattering observed in metals with strong electron correlations.[1–6,8] The quadratic temperature dependence of type (7) also clearly manifests itself in the behavior of the electrical conductivity ρ_a of the samples B1 and B2 when the temperatures are not too high, so that the carrier concentration is essentially constant (see the inset in Fig. 6). At low temperatures the electrical resistivity and the Hall effect in

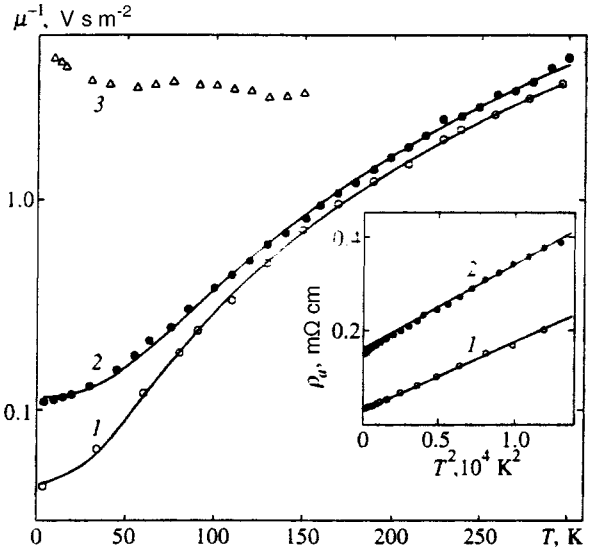

FIG. 6. Temperature curves for the reciprocal mobility $\mu^{-1}=\rho_a/R_{ac}$ in Bi_2Te_3 (the inset shows the electrical resistivity ρ_a as a function of T^2); Curve *1* — the initial sample B1; curves *2* and *3* — the irradiated samples B2 and B4, respectively. △, ●, and ○ represent the experimental data, and the curves are fitting curves (see Eq. (7)).

the samples being studied is weakly temperature-dependent (see Figs. 1 and 2), which justifies using Eqs. (3)–(5) for the degenerate case.

Thus, the bombardment of Bi_2Te_3 crystals by 5-MeV electrons leads to a transition from the *p*-type metallic state to the metallic state with electron conduction. The annealing of defects shifts the chemical potential level from the edge of the conduction band to the top of the valence band through intermediate states in the middle of the energy gap in the electron spectrum, where conduction is by low-mobility carriers. The change in the anisotropy parameter reflects the contribution of different electron bands to conduction along the *c* axis and in the plane, depending on the position of the Fermi level. Here the behavior of the transport properties is similar to that of the ''anomalous'' properties of high-T_c superconducting compounds, observed in a narrow doping region near the metal–insulator transition.[1,2,8] In contrast to doping, electron bombardment leads to reversible changes in the crystal lattice and the electron spectrum, changes that can be reversed by annealing, so that in the irradiated material the metallic *p*-type conduction is restored and quantum oscillations of the resistivity emerge again.

The present work was supported by the State Scientific and Technical Programs *Physics of Quantum and Wave Processes* (Statistical Physics Project No. VIII-6) and *Current Trends in the Physics of Condensed Media* (Superconductivity Project No. 96-051).

*)E-mail: RDNR@neutron.e-burg.su

[1] S. L. Cooper and K. E. Gray, in *Physical Properties of High-Temperature Superconductors*, D. M. Ginsberg (ed.), World Scientific, Singapore (1994).
[2] B. A. Aleksashin, V. I. Voronin, S. V. Verkhovskiĭ, B. N. Goshchitskiĭ, S. A. Davydov, Yu. I. Zhdanov, A. E. Kar'kin, V. L. Kozhevnikov, A. V. Mirmel'shteĭn, K. N. Mikhalev, M. V. Sadovskiĭ, V. V. Serikov, and S. M. Cheshnitskiĭ, Zh. Éksp. Teor. Fiz. **95**, 678 (1989) [Sov. Phys. JETP **68**, 382 (1989)].
[3] P. W. Anderson, Science **235**, 1196 (1987).
[4] T. R. Chien, D. A. Brawner, Z. Z. Wang, and N. P. Ong, Phys. Rev. B **43**, 6242 (1991).
[5] M. Suzuki, Phys. Rev. B **39**, 2312 (1989).
[6] T. R. Chien, Z. Z. Wang, and N. P. Ong, Phys. Rev. Lett. **67**, 2088 (1991).
[7] I. Zambetaki, Qiming Li, E. N. Economou, and C. M. Soukoulis, Phys. Rev. Lett. **76**, 3614 (1996).
[8] A. E. Kar'kin, S. A. Davydov, A. V. Mirmel'shtein, B. N. Goshchitskiĭ, V. Sadovski, and E. Walker, Sverkhprovodimost': Fiz., Khim., Tekhn. **5**, 2215 (1992) [Supercond., Phys. Chem. Technol. **5**, 2091 (1992)].
[9] J. R. Drabble and R. Wolfe, Proc. Phys. Soc. London, Sect. B **69**, 1101 (1956).
[10] B. M. Askerov, *Electron Transport Phenomena in Semiconductors*, World Scientific, Singapore, 1994; Nauka, Leningrad, 1970, p. 161.
[11] S. A. Azou, V. A. Kul'bachinskiĭ, G. A. Mironova, and S. Ya. Skipidarov, Fiz. Tekh. Poluprovodn. **24**, 283 (1990) [Sov. Phys. Semicond. **24**, 173 (1990)].
[12] V. A. Kutasov and L. N. Luk'yanova, Fiz. Tverd. Tela (Leningrad) **32**, 488 (1990) [Sov. Phys. Solid State **32**, 282 (1990)].
[13] K. Seeger, *Semiconductor Physics*, Springer, New York (1973).
[14] J. R. Drabble, in *Progress in Semiconductors*, Vol. 7, A. T. Gibson and R. E. Burgess (Eds.), Temple Press, London (1963), p. 45.
[15] B. M. Gol'tsman, V. A. Kudinov, and I. A. Smirnov, Bi_2Te_3*-based Semiconductor Thermoelectric Materials* [in Russian], Nauka, Moscow (1972), p. 320.
[16] H. C. Montgomery, J. Appl. Phys. **42**, 2971 (1971).
[17] D. L. Greenaway and G. Harbeke, J. Phys. Chem. Solids **26**, 1585 (1965).
[18] *Solids Under Pressure*, W. Paul and D. Warschauer (Eds.), Mir Publishers, Moscow (1966), p. 205 [English original edition: New York (1963)].
[19] E. Muser and W. Pearson, in *Semiconductor Materials. Problems of the Chemical Bond* [in Russian] (a collection of articles translated into Russian), V. P. Zhuze (ed.), Inostrannaya Literatura, Moscow (1960), pp. 220–248.
[20] P. S. Kireev, *Semiconductor Physics*, Mir Publishers, Moscow (1978).
[21] C. C. Y. Kwan, J. Basinski, and J. C. Woolley, Phys. Status Solidi B **48**, 699 (1971).
[22] L. R. Testardi, P. J. Stiles, and E. Burstain, Solid State Commun. **1**, 28 (1963).
[23] P. Pechur and G. Toussaint, in *Proc. 1st Eur. Conf. Thermoelec., Cardiff, 15–17 Sept., 1987*, London (1988), p. 394.
[24] J. P. Fleurial, S. Y. Rocca, H. Scherrer, and S. Scherrer, in *Proc. 1st Eur. Conf. Thermoelec., Cardiff, 15–17 Sept., 1987*, London (1988), p. 74.
[25] V. A. Kul'bachinskiĭ, V. V. Shchennikov, Ya. Gorak, and P. Loshtyak, Fiz. Tverd. Tela (St. Petersburg) **36**, 526 (1994) [Phys. Solid State **36**, 292 (1994)].

Translated by Eugene Yankovsky

Collapse in the Stoner–Wohlfarth noninteracting-particle model

A. M. Afanas'ev and M. A. Chuev*)

Physicotechnical Institute, Russian Academy of Sciences, 117218 Moscow, Russia

J. Hesse

Institut für Metallphysik und Nukleare Festkörperphysik, Technische Universität Braunschweig, D-38106 Braunschweig, Germany

(Submitted 2 December 1997)
Zh. Éksp. Teor. Fiz. **113**, 1799–1815 (May 1998)

A theory of Mössbauer spectra of noninteracting Stoner–Wohlfarth (SW) particles interacting with rf magnetic fields is developed. The theory makes it possible to calculate the absorption spectra for arbitrary frequency and amplitude of the rf field. The main features of the Stoner–Wohlfarth model are discussed. The Liouville superoperator formalism is used to generalize the results to the case of arbitrarily time-varying hyperfine fields at a nucleus. To understand the qualitative features of the collapse effect that are observed in the Mössbauer spectra of SW particles the particular case of a circularly polarized hyperfine field is studied, and an analytical expression is obtained describing the Mössbauer spectra for this case. An analysis is also made for weak rf magnetic fields and in this case the resonance behavior of the Mössbauer lines is traced as a function of the frequency of the rf field. © *1998 American Institute of Physics.* [S1063-7761(98)01705-3]

1. INTRODUCTION

The method of investigating Mössbauer spectra of the hyperfine structure by irradiating the sample with a radio-frequency (rf) field is now attracting increasing attention. This method makes it possible to study the relaxation properties of magnetic alloys at frequencies of the order of hundreds of MHz, which is of great practical interest. Numerous experimental works, where the diverse behavior of the shapes of the Mössbauer spectra as a function of the frequency and amplitude of the rf field was demonstrated qualitatively, have been published in connection with this problem.[1–11] As the amplitude of the rf field increases, the hyperstructure, which is well resolved in the absence of the rf field, collapses to a single line. This phenomenon, discovered by Pfeiffer in 1971, is called the collapse effect.

Theoretical investigations in this field are actually only just beginning. The theoretical work published thus far analyzes only the case of extremely strong rf fields and does not contain even a rough description for the case of rf fields of arbitrary intensity.[12–14] Only very recently we were the first to propose a theory that made it possible to calculate in a model the Mössbauer spectra for arbitrary frequencies and amplitudes of the rf field.[15]

In the latest work the magnetic alloy was treated, by analogy to the Stoner–Wohlfarth (SW) model,[16] as a collection of small uniformly magnetized particles with randomly oriented axes of easiest magnetization. The rf-field-induced magnetization reversal process was treated as random jumps from one orientation of the axis of easiest magnetization to the opposite orientation. This assumption greatly simplified the analysis and made it possible to obtain a closed-form expression for the Mössbauer absorption spectrum. Even in this very simple case the final result had a very nontrivial form and consisted of a double integral of an operator function. Nonetheless, the formulas obtained can be programmed on a computer, and in Ref. 15 a number of specific examples were studied (see Ref. 15 for details).

It is perfectly obvious that the model developed in Ref. 15 is not general. For example, in the Stoner–Wohlfarth model[16] magnetization reversal of a particle under the action of a magnetic field is accompanied by a deflection of the magnetic moment of the particle away from the axis of easiest magnetization. This circumstance makes it much more difficult to calculate the Mössbauer absorption spectra and makes it necessary to use more refined mathematical methods of analysis. Nonetheless, if the SW model, where the magnetic relaxation admits a relatively simple mathematical description, is strictly followed, it is possible to obtain even in this case closed-form expressions for the Mössbauer absorption spectra and to perform concrete calculations for arbitrary rf amplitudes and frequencies. The present paper is devoted to this question. In what follows we shall work strictly in the SW model in its original form,[16] where interparticle interactions were neglected.

A detailed derivation of the formulas describing the Mössbauer absorption spectra in the case when the time-dependent hyperfine field at a nucleus reverses direction, i.e., it remains parallel to the same axis, is given in Sec. 2. The Stoner–Wohlfarth model is briefly discussed in Sec. 3. In Sec. 4 the results of Sec. 2 are extended to the case of an arbitrary time-dependent hyperfine magnetic field by replacing the Hamiltonian operators by Liouville superoperators. To understand the characteristic behavior of the Mössbauer spectra in the SW model, the case of a circularly polarized hyperfine field is specially studied in Sec. 5. This example admits a simple analytical description and makes it possible to show a number of new features of the collapse effect

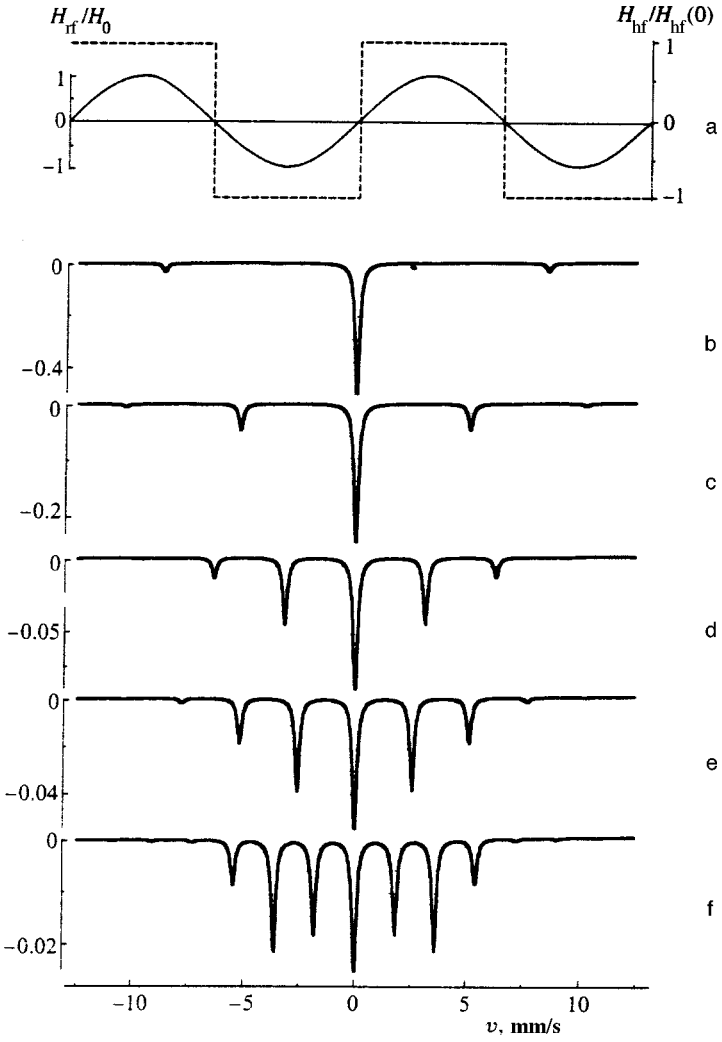

FIG. 1. a—Time dependence of the hyperfine magnetic field (dashed line) in a strong rf field (solid line) and the corresponding Mössbauer spectra for rf frequencies $\omega_{rf}=100$, 50, 36.8, 30, and 21.1 MHz (b–f). Here and below $\omega_L=50$ MHz.

which are manifested in the Mössbauer absorption spectra in the SW model. The case of weak rf fields, where the general expressions greatly simplify and the resonant character of the rf frequency dependence of the Mössbauer spectra can be precisely traced, is analyzed in Sec. 6.

2. MÖSSBAUER SPECTRA OF THE HYPERFINE STRUCTURE UNDER THE ACTION OF AN RF FIELD IN THE ADIABATIC APPROXIMATION: DETERMINISTIC DESCRIPTION OF THE HYPERFINE FIELD

A rf field influences the hyperfine spectra via the hyperfine field $\mathbf{H}_{hf}(t)$ acting on a nucleus. Under the action of the rf field the hyperfine field becomes time-dependent as result of complicated relaxation processes. The simplest and physically clearest situation occurs in the cases when the relaxation processes are rapid enough that $\mathbf{H}_{hf}(t)$ follows some macroscopic characteristics of the sample—in the SW model the magnetization $\mathbf{M}_i(t)$ of a particle, where $\mathbf{M}_i(t)$ in turn varies in time as a result of the action of the rf field. If a model of magnetization reversal is given, then $\mathbf{M}_i(t)$ and therefore also $\mathbf{H}_{hf}(t)$ are known functions and the Hamiltonian of the system, which includes the hyperfine interaction and the interaction of a γ-ray with a nucleus, will have the following form for the ground and excited states of a nucleus:

$$\hat{H}=\hat{H}_0+g_{g,e}\mu_N\hat{\mathbf{I}}^{(g,e)}\mathbf{H}_{hf}(t)+\hat{V}_{\gamma N}(t_0), \tag{1}$$

where $\hat{H}_0$ is the nuclear Hamiltonian determining the energy levels of a nucleus in the ground and excited states, neglecting the hyperfine interaction, and the second term in Eq. (1) describes the hyperfine interaction of a nucleus. Here μ_N is the nuclear magneton; $g_{g,e}$ is the nuclear g factor; $\hat{\mathbf{I}}^{(g,e)}$ are the nuclear spin for the ground (g) and excited (e) states of a nucleus; and, the operator $\hat{V}_{\gamma N}(t_0)$ describes the interaction of a γ-ray with a nucleus, t_0 is being the time at which this interaction is switched on. According to the general theory of the resonant interaction of radiation with matter,[17] we take $V(t)=0$ for $t<t_0$. As one can see from Eq. (1), the form of the absorption spectrum will be determined through the dependence $\mathbf{H}_{hf}(t)$.

In strong rf fields, most of the time each particle will possess a direction of magnetization that is essentially parallel to the direction of the rf field and, assuming extremely rapid relaxation, its direction will change instantaneously

when the direction of the rf field changes (see Fig. 1a). If the perpendicular components of the hyperfield are neglected, then we have for the wave functions of the nucleus in the ground and excited states, $\psi_m(t)$ and $\psi_M(t)$ respectively,

$$\psi_m(t,t_0)=\exp\left[-ig_g\mu_N m\int_{t_0}^{t}H_{\rm hf}(t')dt'\right]|m\rangle,$$

$$\psi_M(t,t_0)=\exp\left[-ig_e\mu_N M\int_{t_0}^{t}H_{\rm hf}(t')dt'+\left(i\frac{E_0}{\hbar}-\frac{\Gamma_0}{2}\right)(t-t_0)\right]|M\rangle. \tag{2}$$

Here m and M are the projections of the nuclear spin on the direction of the hyperfine field in the ground and excited states, respectively, and E_0 and Γ_0 are the energy and width of the excited state of the nucleus.

According to Ref. 17, we have for the amplitude of the absorption of a γ-ray with energy $E=\hbar\omega$

$$c_{mM}(\omega)=\int_{t_0}^{\infty}\langle\psi_M^*(t)|\hat{V}_{\gamma N}(t_0)\exp[i\omega(t-t_0)]|\psi_m(t)\rangle dt \equiv V_{mM}\int_{t_0}^{\infty}\exp\left[\int_{t_0}^{t}i\omega_{mM}(t')dt'+i\tilde{\omega}(t-t_0)\right]dt, \tag{3}$$

where $\omega_{mM}(t)=(g_eM-g_gm)\mu_N H_{\rm hf}(t)$, $\tilde{\omega}=\omega+i\Gamma_0/2$, and V_{mM} are the matrix elements of the nuclear current operator, which also includes the probability of the Mössbauer effect. The square of the absolute value of the absorption amplitude determines the absorption cross section.[17] It is also necessary to average over the switch-on time t_0. Ordinarily, no special attention is given to t_0, but in this case there is a time-dependent rf field, which determines its own reference time. Since the rf field is periodic, it is sufficient to average over one period. On this basis we have for the absorption cross section

$$\sigma(\omega)=\sigma_a\sum_{\alpha}|C_\alpha|^2\varphi_\alpha(\omega), \tag{4}$$

where

$$\varphi_\alpha(\omega)=\frac{1}{T_{\rm rf}}\int_0^{T_{\rm rf}}\left|\int_{t_0}^{\infty}\exp\left\{\int_{t_0}^{t}i[\tilde{\omega}-\omega_\alpha(t')]dt'\right\}dt\right|^2dt_0 \equiv\frac{1}{T_{\rm rf}}\int_0^{T_{\rm rf}}dt_0\int_{t_0}^{\infty}dt_1\,\exp[-\Gamma_0(t_1-t_0)] \times\int_{t_1}^{\infty}dt\,\exp\left\{\int_{t_1}^{t}i[\tilde{\omega}-\omega_\alpha(t')]dt'\right\}+\text{c.c.} \tag{5}$$

Here $\sigma_a=f_a n_a s_0 t_a$ is the effective thickness of the absorber, f_a is the probability of the Mössbauer effect in the absorber, n_a is the density of resonant nuclei in the absorber, σ_0 is the transverse absorption cross section, t_a is the thickness of the absorber, $\alpha\equiv m$ and M, the coefficients C_α determine the intensity of the corresponding transitions and can be expressed in terms of the Clebsch–Gordan coefficients, $T_{\rm rf}=2\pi/\omega_{\rm rf}$ is the period of the rf field, and $\omega_{\rm rf}$ is the frequency of the rf field.

The periodicity of the rf field makes it possible to simplify the expression (5). The integral over t_0 can be performed by parts, the result being

$$\varphi_\alpha(\omega)=\frac{1}{\Gamma_0 T_{\rm rf}}\int_0^{T_{\rm rf}}dt_1\int_{t_1}^{\infty}dt\times\exp\left\{\int_{t_1}^{t}i[\tilde{\omega}-\omega_\alpha(t')]dt'\right\}+\text{c.c.} \tag{6}$$

Since $\omega_\alpha(t)=\omega_\alpha(t+T_{\rm rf})$, we easily find

$$\varphi_\alpha(\omega)=\frac{1}{\Gamma_0 T_{\rm rf}}\int_0^{T_{\rm rf}}dt_1\int_{t_1}^{t_1+T_{\rm rf}}dt\times\frac{\exp[i\tilde{\omega}(t-t_1)]}{1-\exp(i\tilde{\omega}T_{\rm rf})\hat{G}(t_1,t_1+T_{\rm rf})}G(t_1,t)+\text{c.c.}, \tag{7}$$

where

$$G(t_1,t)=\exp\left\{\int_{t_1}^{t}dt'(-i\omega_\alpha(t))\right\}. \tag{8}$$

The function $G(t_1,t)$ can be factorized as

$$G(t_1,t)=G^{-1}(0,t_1)G(0,t). \tag{9}$$

As a result, Eq. (7) acquires the simpler form

$$\varphi_\alpha(\omega)=\frac{1}{\Gamma_0 T_{\rm rf}}\int_0^{T_{\rm rf}}dt_1\int_{t_1}^{t_1+T_{\rm rf}}dt\;G^{-1}(0,t_1)\times\frac{\exp[i\tilde{\omega}(t-t_1)]}{1-\exp(i\tilde{\omega}T_{\rm rf})\hat{G}(0,T_{\rm rf})}G(0,t)+\text{c.c.} \tag{10}$$

This form, where the integration limits are finite, is most convenient for specific calculations for an arbitrary periodic time-dependence of the hyperfine field $\mathbf{H}_{\rm hf}(t)$.

The integral (10) can be easily calculated in the case when $\mathbf{H}_{\rm hf}$ varies in time as shown in Fig. 1a. The result is

$$\varphi_\alpha(\omega)=\frac{1}{8\Gamma_0}\left\{\frac{1+\exp(i\tilde{\omega}T_{\rm rf})}{1-\exp(i\tilde{\omega}T_{\rm rf})}[F_\alpha(\tilde{\omega})+F_\alpha(-\tilde{\omega})]+[F_\alpha(\tilde{\omega})-F_\alpha(-\tilde{\omega})]\right\}+\text{c.c.}, \tag{11}$$

where

$$F_\alpha(\pm\tilde{\omega})=G^{\pm}(\tilde{\omega},\omega_\alpha)+G^{\pm}(\tilde{\omega},-\omega_\alpha),$$

$$G^{\pm}(\tilde{\omega},\omega_\alpha)=(\tilde{y}/y_\alpha+2)f^{\pm}(\tilde{y}+y_\alpha),$$

$$f^{+}(x)=\frac{1-\cos x}{x^2},\quad f^{-}(x)=i\frac{x-\sin x}{x^2},$$

$$\tilde{y}=\frac{\tilde{\omega}T_{\rm rf}}{2},\quad y_\alpha=\frac{\omega_\alpha T_{\rm rf}}{2}.$$

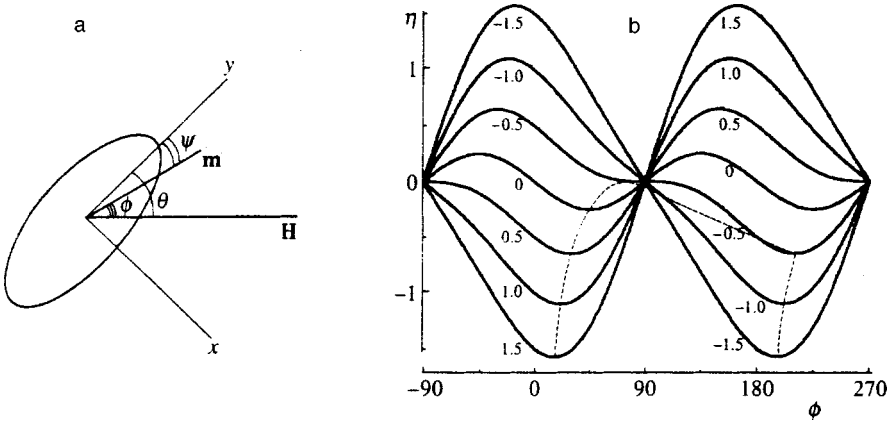

FIG. 2. a—Magnetization **m** of a SW particle in an external magnetic field **H**. b—Normalized energy $\eta=E/(N_x-N_y)VM_0^2$ of a SW particle ($\theta=45°$) versus the angle ϕ for different values of the normalized external magnetic field $h=H(N_x-N_y)VM_0$; the dashed line shows the variation of the angle ϕ with decreasing h from a positive value above h_c to negative values.

Examples of the transformation of the hyperfine structure spectra as a function of the ratio of the rf and the Larmor frequency are presented in Figs. 1b–f. In all cases the spectra consist of a single central line and satellites. According to Eq. (11), we have for the intensities of the satellites in the case $\omega_{\rm rf}\gg\Gamma_0$

$$I_k=2\sigma_a\sum_\alpha\left[\frac{y_\alpha}{(\pi k)^2-y_\alpha^2}\right]^2[1-(-1)^k\cos y_\alpha]. \tag{12}$$

These results are in complete agreement with Refs. 13 and 14.

3. STONER–WOHLFARTH MODEL

The model proposed by Stoner and Wohlfarth in 1946 is widely used to describe magnetization reversal in magnetic alloys.[16] In this model a ferromagnet is treated as a collection of individual particles or clusters, inside which a strong exchange interaction operates, so that each particle can be treated as uniformly magnetized with magnetization M_0. For simplicity it is assumed that each particle is a prolate ellipsoid of revolution with a random orientation of the axes in space (see Fig. 2a). As a consequence of the nonsphericity of the particles, there arises a shape energy

$$E_d=\frac{VM_0^2}{2}(N_y\cos^2\psi+N_x\sin^2\psi), \tag{13}$$

where V is the particle volume and N_x and N_y are demagnetization factors along the x and y axes, respectively.

The SW model also assumes that

$$E_d\gg kT, \tag{14}$$

so that in the absence of external fields the magnetic moment of each particle occupies one of two positions in the direction of the axis of easiest magnetization. On account of the relation (14) jumps between these two states are assumed to be very slow so that they can be neglected during the measurement time.

If an external magnetic field H acts on a particle, then the total Hamiltonian of the particle is

$$\hat{H}=-HVM_0\cos\phi+\frac{VM_0^2}{2}(N_y\cos^2\psi+N_x\sin^2\psi). \tag{15}$$

Here the first term describes the energy of the particle in the external magnetic field. Figure 2b shows a series of curves of the total energy of a particle as a function of the direction of the magnetization vector for different values of the normalized magnetic field intensity

$$h=\frac{H}{(N_x-N_y)VM_0}. \tag{16}$$

As one can see from Fig. 2b, when an external magnetic field is applied, the energy minima shift as a function of the angle ϕ, i.e., the magnetic moment tilts away from the axis of easiest magnetization (see Fig. 2a). For weak magnetic fields $h<h_c$ two energy minima of different depth and two maxima remain on the curve. If the external field is stronger than h_c, we have curves with one minimum (h_c is the critical magnetic field intensity, which depends on the orientation of the particle[16]).

An important assumption of the SW model is that there exist fast relaxation processes such that as the external field varies, the magnetic moment of a particle follows the position of and instantaneously changes direction in accordance with the local minimum and passes instantaneously into the absolute minimum only for fields $h>h_c$. The dashed line in Fig. 2b shows the variation of the direction of the magnetic moment as a function of h.

The behavior of the magnetic moment of a particle as a function of the external magnetic field H can be calculated according to the scheme described above. The corresponding curves for an external periodic field

$$H(t)=H_0\sin\omega t \tag{17}$$

are displayed in Fig. 3, which displays the projections of the magnetic moment on the direction of the magnetic field for different groups of particles with different orientations of the axes of easiest magnetization, as well as the general magnetization curve with all particles randomly oriented. As one can see from this figure, for fields above h_c the magnetization curve reaches a stationary closed curve irrespective of the initial position of the magnetic moment of the particle at $H=0$. For weak fields memory of the initial state is retained, and for each particle there are two magnetization curves, each curve corresponding to a different initial state, distinguished by an opposite orientation of the magnetic moment.

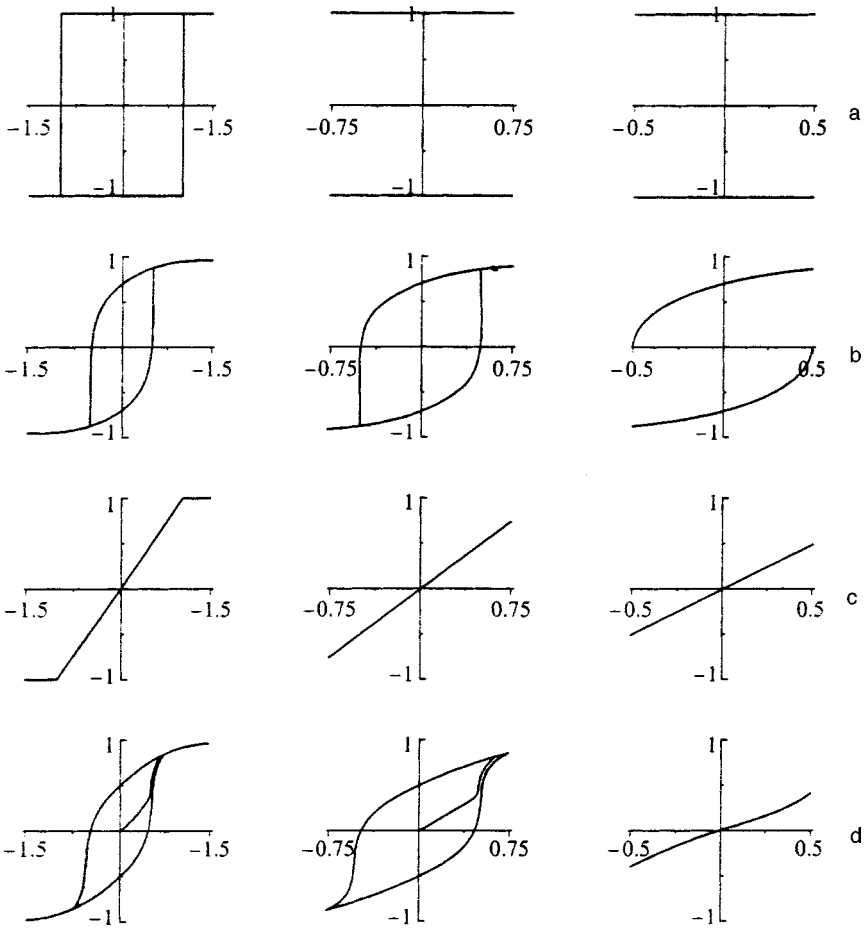

FIG. 3. Magnetization curves for SW particles with angles $\theta=0°$, 45°, and 90° (a–c) and for an ensemble of randomly oriented SW particles (*d*) for different rf amplitudes $h_0=1.5$, 0.75, and 0.5 (left to right).

Since the critical field h_c depends on the orientation of the axis of easiest magnetization of a particle relative to the direction of the magnetic field, for certain amplitudes of the external field H_0 in the range $0.5<h_0<1$ (see the series of curves in Figs. 3a–c) the magnetization of some particles will reverse in the strong-field regime and that of other particles will reverse in the weak-field regime. In the present model these features of the magnetization reversal process will be important for the formation of the hyperfine structure of the Mössbauer spectra.

We note that nontrivial role of relaxation in this model. On the one hand the relaxation process is assumed to be rapid enough for the magnetic moment to follow the position of the local energy minimum, while on the other hand this process must be quite slow compared with the jumps between local minima. It is obvious that both conditions can be realized only if the condition (14) holds with a large margin, so that the rate of above-barrier transitions between local minima can be assumed to be zero with exponential accuracy. The condition (14) is well-known to hold for large particles. As the particle volume decreases, the role of relaxation processes changes substantially and it becomes necessary to take into account jumps between local minima. This extension of the SW model can be accomplished on the basis of the approach proposed in Ref. 15.

4. MÖSSBAUER SPECTRA FOR STONER–WOHLFARTH PARTICLES

Thus, in the Stoner–Wohlfarth model the magnetic moment of the particles changes direction and a special investigation is required in order to take this circumstance into account in the Mössbauer absorption spectra. Such an investigation is much more complicated than not only the case examined in Sec. 2 but also the more general case studied in Ref. 15. In the present case the hyperfine interaction is determined by the Hamiltonian (1) without any restrictions, i.e., $\mathbf{H}_{\rm hf}(t)$ can change direction arbitrarily. Since the hyperfine interaction operators taken at different times do not commute, a simple solution of the kind obtained in Sec. 2 no longer exists.

In this case the wave functions for the ground and excited states are sought in the form

$$|\psi^{(g)}(t,t_0)\rangle=c_{mm_0}(t,t_0)|m\rangle,$$
$$|\psi^{(e)}(t,t_0)\rangle=\tilde{c}_{MM_0}(t,t_0)|M\rangle. \tag{18}$$

Here m, M and m_0, M_0 are the projections of the nuclear spin on a distinguished axis at the times t and t_0, respectively. In the expressions (18) and below summation over repeated indices is assumed. From the Schrödinger equations for the coefficients $c_{mm_0}(t,t_0)$ and $\tilde{c}_{M_0M}(t,t_0)$ we easily find

$$c_{mm_0}(t,t_0)=\langle m|\tilde{T}\ \exp\Bigg[-ig_g\mu_N m \times\int_{t_0}^{t}\hat{H}_{\rm hf}(t')dt'\Bigg]|m_0\rangle, \tag{19}$$

$$\tilde{c}_{MM_0}(t,t_0)=\langle M|\tilde{T}\ \exp\Bigg[-ig_e\mu_N M\int_{t_0}^{t}\hat{H}_{\rm hf}(t')dt' +\left(i\frac{E_0}{\hbar}-\frac{\Gamma_0}{2}\right)(t-t_0)\Bigg]|M_0\rangle.$$

Taking account of the averaging with respect to the initial states over a period of the rf field, the absorption cross section for particles with a fixed orientation (i) of the axis of easiest magnetization will be given by the expression

$$\sigma_i(\omega)=\frac{1}{T_{\rm rf}}\int_0^{T_{\rm rf}}\sum_{m_0,M_0}\left|\int_{t_0}^{\infty}c^{*}_{M_0M}(t,t_0) \times V_{Mm}c_{mm_0}(t,t_0)e^{i\omega(t-t_0)}dt\right|^2 dt_0. \tag{20}$$

To transform this expression it is convenient to introduce the Liouville operator $\hat{\mathbf{L}}_{\hat{H}}(t)$, whose action on the ordinary operator $\hat{A}(t)$ is determined by the relation[18]

$$i\frac{\partial\hat{A}(t)}{\partial t}=\hat{\mathbf{L}}_{\hat{H}}(t)\hat{A}(t)=\frac{1}{\hbar}[\hat{H}(t),\hat{A}(t)]. \tag{21}$$

In our case the superoperator $\hat{\mathbf{L}}_{\hat{H}}(t)$ operates in a space with $(2I^{(g)}+1)(2I^{(e)}+1)$ variables, and for the case of ^{57}Fe nuclei it is a matrix of rank 8. With the aid of this operator the expression (20) can be put into a form similar to the left-hand side of Eq. (5):

$$\sigma_i(\omega)=\frac{1}{T_{\rm rf}}\int_0^{T_{\rm rf}}dt_0\int_{t_0}^{\infty}dt_1\ \exp[-\Gamma_0(t_1-t_0)]\int_{t_1}^{\infty}{\rm Tr} \times\left\{\hat{V}\left[\hat{T}\ \exp\left\{\int_{t_1}^{t}i[\tilde{\omega}\hat{\mathbf{I}}-\hat{\mathbf{L}}_{\hat{H}}(t')dt']\right\}\hat{V}\right]\right\}dt+{\rm c.c.} \tag{22}$$

Averaging over t_0 is performed completely analogously to Sec. 2. This gives

$$\sigma_i(\omega)=\frac{1}{\Gamma_0 T_{\rm rf}}\int_0^{T_{\rm rf}}dt_1\int_{t_1}^{\infty}{\rm Tr}\left\{\hat{V}\left[\hat{T} \times\exp\left\{\int_{t_1}^{t}i[\tilde{\omega}\hat{\mathbf{I}}-\hat{\mathbf{L}}_{\hat{H}}(t')]dt'\right\}\right]\hat{V}\right\}dt+{\rm c.c.}, \tag{23}$$

and since the rf field is periodic, the expression (23) can be put into the form

$$\sigma_i(\omega)=\frac{1}{\Gamma_0 T_{\rm rf}}\sum_{\pi}\int_0^{T_{\rm rf}}dt_1\int_{t_1}^{t_1+T_{\rm rf}}dt\ {\rm Tr} \times\left[\hat{V}_\pi\hat{\mathbf{G}}^{-1}(0,t_1)\frac{\exp[i\tilde{\omega}(t-t_1)]}{\hat{\mathbf{I}}-\exp(i\tilde{\omega}T_{\rm rf})\hat{\mathbf{G}}(0,T_{\rm rf})} \times\hat{\mathbf{G}}(0,t)\hat{V}_\pi\right]+{\rm c.c.}, \tag{24}$$

where

$$\hat{\mathbf{G}}(0,t)=\hat{T}\ \exp\left\{\int_0^t dt'[-i\hat{\mathbf{L}}_{\hat{H}}(t')]\right\} \tag{25}$$

($\hat{T}$ is the chronological ordering operator). In Eq. (24) summation is also performed over different polarizations π of the incident γ-rays. The total absorption cross section is obviously determined by averaging over all the differently oriented particles.

In accordance with the definition of the $\hat{T}$ product the operator $\hat{\mathbf{G}}(0,t)$ can be represented as

$$\hat{\mathbf{G}}(0,t)=\lim_{n\to\infty}\prod_{k=1}^{n}\ \exp(-i\hat{\mathbf{L}}_{\hat{H}}(t_k)\Delta t), \tag{26}$$

where $\Delta t=t/n$ and $t_k=t-k\Delta t$. If the magnetic dynamics of the system is known, i.e., $\mathbf{M}_i(t)$, and therefore $\mathbf{H}_{\rm hf}^{(i)}(t)$ are known functions (see Fig. 3), it is not difficult to calculate the matrix elements of the operator $\hat{\mathbf{G}}(0,t)$ numerically in analogy to ordinary integrals. Then the Mössbauer absorption spectra can be calculated directly using Eqs. (24)–(26). We note that the calculations are much more involved than in the preceding case, where the field varied only in magnitude and not in direction. The computational results for the Mössbauer absorption spectra for different rf amplitudes and frequencies are displayed in Figs. 4 and 5.

As one can see from Fig. 4, as the rf amplitude increases, in strong magnetic rf fields the hyperfine structure that is well-resolved in weak rf fields collapses to a single central line with satellites located at distances $\pm n\omega_{\rm rf}$, where n is an integer. As the rf frequency decreases, the intensity of the satellites increases. In intermediate fields a much more complicated picture of the transformation of the spectra is observed than in the case of the hyperfine structure spectra calculated in Ref. 15, where the perpendicular components of the hyperfine splitting were assumed to be zero.

Figure 5 shows a series of partial spectra for particles with different orientations with angles $\theta=0°$, 45°, and 90°, for which the role of the perpendicular components of $\mathbf{H}_{\rm hf}(t)$ is most conspicuous. It is easy to see that spectra with $\theta=0°$ and 90° are qualitatively different from one another, and for $\theta=90°$ a unique hyperfine structure appears for both the central line and the satellites. Moreover, splitting of the individual components of the hyperfine structure is observed (Figs. 5b, and 5c) for weak fields and rf frequency equal to the frequency of the hyperfine splitting of the nuclear ground

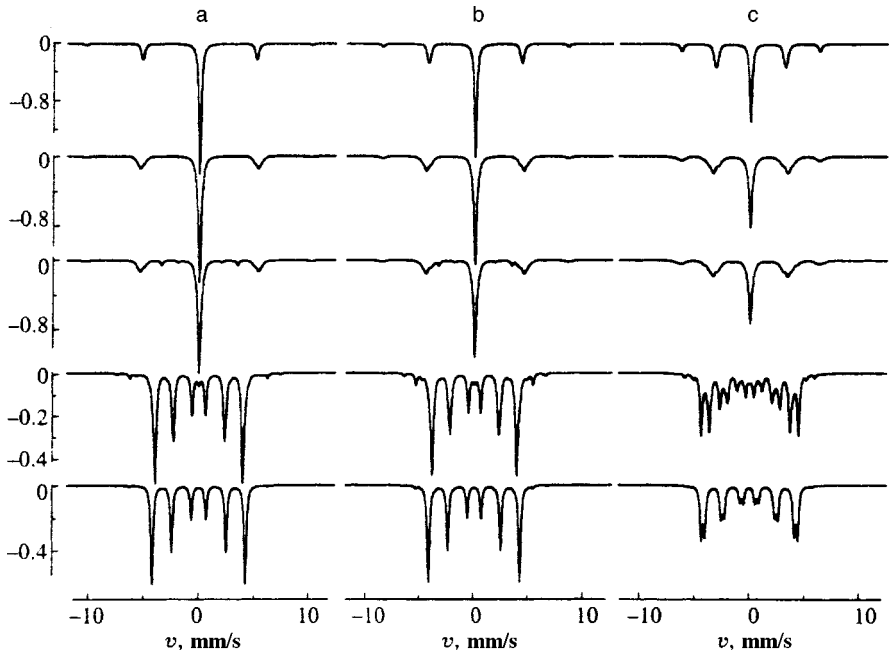

FIG. 4. Computed Mössbauer absorption spectra for an ensemble of randomly oriented SW particles in a rf field with different frequencies $\omega_{rf}=60$, 50, and 36.8 MHz (a–c) and amplitudes $h_0=5$, 1, 0.75, 0.5, and 0.25 (top to bottom).

state. These features occur over a quite wide range of angles θ and they even appear clearly in the total spectra (Fig. 4). These effects require closer attention and will be discussed below.

5. MÖSSBAUER SPECTRA FOR A CIRCULARLY POLARIZED HYPERFINE FIELD

It is easy to see that for $h_0=1$ and $\theta=90°$ the hyperfine field at a nucleus is strictly circularly polarized, i.e.,

$$\mathbf{H}_{hf}(t)=\mathbf{n}_x\cos(\omega_{rf}t)+\mathbf{n}_y\sin(\omega_{rf}t). \tag{27}$$

As we shall see below, the expression for the Mössbauer absorption spectrum has a relatively simple analytical solution in this case.

Let us consider the wave function of the nuclear ground state. It is convenient to seek the solution in the form

$$|\psi^{(g)}(t)\rangle=\hat{T}\exp\left(-i\int_{t_0}^{t}\hat{H}^{(g)}(t')dt'\right)|\psi^{(g)}(t_0)\rangle, \tag{28}$$

where

$$\hat{H}^{(g)}(t)=\omega_g\hat{\mathbf{I}}\cdot\mathbf{H}_{hf}(t). \tag{29}$$

Here ω_g is the hyperfine splitting constant for the ground state. A simpler expression can be obtained for $|\psi^{(g)}(t)\rangle$ by means of a unitary transformation into a coordinate system rotating with frequency ω_{rf} around the z axis:

$$\begin{aligned}|\psi^{(g)}(t)\rangle&=\exp(i\omega_{rf}\hat{I}_z t_0)\exp(-i\omega_{rf}\hat{I}_z t_0)\hat{T}\\&\times\exp\left(-i\int_{t_0}^{t}\hat{H}^{(g)}(t')dt'\right)\\&\times\exp(i\omega_{rf}\hat{I}_z t)\exp(-i\omega_{rf}\hat{I}_z t)|\psi^{(g)}(t_0)\rangle\\&\equiv\exp(i\omega_{rf}\hat{I}_z t_0)\exp(-i\tilde{\hat{H}}^{(g)}(t-t_0))\\&\times\exp(-i\omega_{rf}\hat{I}_z t)|\psi^{(g)}(t_0)\rangle,\end{aligned} \tag{30}$$

where

$$\tilde{\hat{H}}^{(g)}=\omega_{rf}\hat{I}_z+\omega_g\hat{I}_x \tag{31}$$

is a time-independent operator.

We now introduce the eigenfunctions $|\tilde{m}\rangle$ and $|m\rangle$ and the eigenvalues $\omega_{\tilde{m}}$ and ω_m for the operators $\tilde{\hat{H}}^{(g)}$ and $\omega_{rf}\hat{I}_z$. It is easy to see that

$$\omega_{\tilde{m}}=\sqrt{\omega_{rf}^2+\omega_g^2}\,\tilde{m},\quad \omega_m=\omega_{rf}m, \tag{32}$$

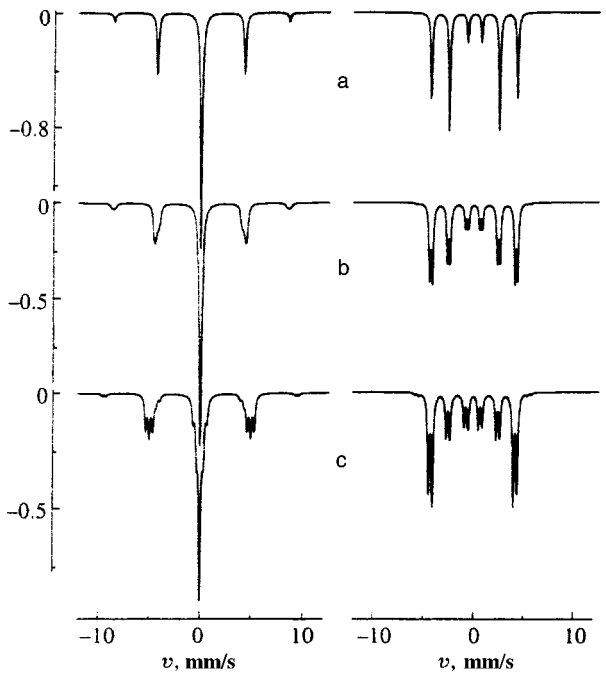

FIG. 5. Computed Mössbauer spectra for SW particles with $\theta=0°$, 45°, and 90° (a–c) in a strong rf field, $h_0=1$, with $\omega_{rf}=50$ MHz (left side) and in a weak rf field, $h_0=0.25$, at the resonance frequency $\omega_{rf}=\omega_g=36.8$ MHz (right side).

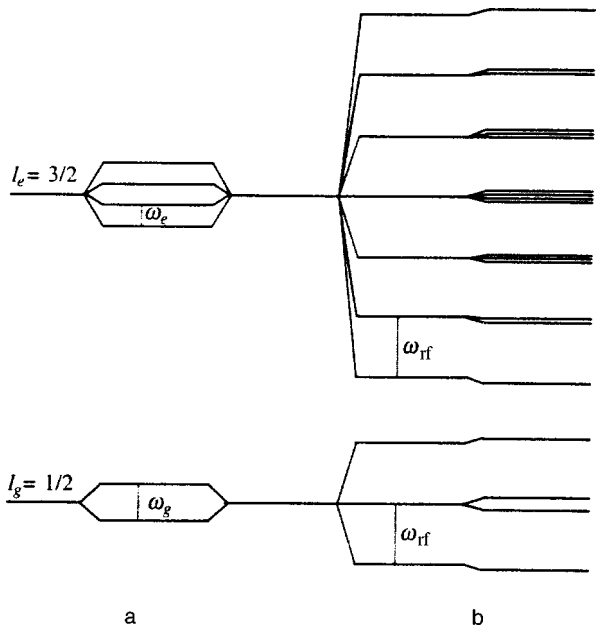

FIG. 6. Scheme of the splitting of the energy levels of a nucleus in excited and ground states in a static hyperfine field (a) and in a circularly polarized hyperfine field $\mathbf{H}_{\rm hf}(t)$ ($\theta=90°$, $h_0=1$) (b).

where m is the projection of the nuclear spin on the z axis and $\tilde{m}$ is the projection of the nuclear spin on the quantization axis for which the operator $\hat{\tilde{H}}^{(g)}$ is diagonal. Using these expressions it is easy to express the function $c_{mm_0}(t,t_0)$ as

$$c_{mm_0}(t,t_0)=\langle m_0|\tilde{m}\rangle\langle\tilde{m}|m\rangle\exp(i(\omega_{m_0}-\omega_{\tilde{m}})t_0)\times\exp(i(\omega_{\tilde{m}}-\omega_m)t). \quad (33)$$

Similarly, we find for the excited state

$$c_{MM_0}(t,t_0)=\langle M_0|\tilde{M}\rangle\langle\tilde{M}|M\rangle\exp(i(\omega_{M_0}-\omega_{\tilde{M}})t_0)\times\exp(i(\omega_{\tilde{M}}-\omega_M)t)\times\exp(i(E_0/\hbar+i\Gamma/2)(t-t_0)). \quad (34)$$

As one can see from Eqs. (33) and (34), a nucleus with spin I behaves under the action of an external circularly polarized field similarly to a system with the energy levels (quasilevels)

$$\omega_i=\omega_{\tilde{M}}-\omega_M. \quad (35)$$

The number of these quasilevels is equal to $(2I+1)^2$. The corresponding scheme of the splitting of the energy levels of the ground and excited states of the ^{57}Fe nucleus is shown in Fig. 6. It is obvious that these levels will appear in the absorption spectra as sharp lines with natural width, which we in fact observe in Fig. 5c. Using Eqs. (33) and (34) it is not difficult to obtain analytical expressions for the absorption spectrum. In so doing, it is convenient to employ not the final result (24) but rather Eq. (23). Simple manipulations yield

$$\sigma_i(\omega)=\frac{1}{\Gamma_0}\sum_{\pi}\sum_{\substack{mM\\ \tilde{m}\tilde{M}}} V^{(\pi)+}_{\tilde{m}\tilde{M}}\times\frac{\langle\tilde{m}|m\rangle\langle M|\tilde{M}\rangle}{i(\tilde{\omega}-\omega_{\tilde{M}}+\omega_{\tilde{m}}+\omega_M-\omega_m)}V^{(\pi)}_{Mm}+\text{c.c.} \quad (36)$$

The calculations performed using Eq. (36) and Eqs. (24)–(26) are in complete agreement with one another.

Thus, in the case studied above the absorption spectrum consists of sharp lines with a Lorentzian shape analogous to the case of a static hyperfine field but with a much larger number of lines. In the general case, for the transition studied the spectrum consists of 64 lines. The selection rules dictated by the multipole nature of the $M1$ transition make transitions with indices $M=\pm 3/2\rightarrow m=\mp 1/2$ forbidden and the number of allowed lines is thereby reduced to 48, only 24 doubly-degenerate lines being actually observed. Moreover, for large $\omega_{\rm rf}$ the lines separate into a central group consisting of six lines and side groups (satellites) with a resolved hyperfine structure with 5, 3, and 1 lines with increasing distance from center.

As direct calculations show, this picture of the more or less clearly expressed hyperfine structure of both the central line and the satellites remains both in a quite wide range of angles θ near $\theta=90°$ and in a wide range of frequencies of the rf field.

6. WEAK RF-FIELD LIMIT

The case of weak rf fields is also of special interest, since the typical resonant effects in the form of splitting of the hyperfine lines arise even for very weak fields (see Fig. 5c). If the rf amplitude for each particle is much less than the anisotropy energy ($h_0\ll 1$), then the magnetic moment of a particle and therefore $\mathbf{H}_{\rm hf}(t)$ also execute small oscillations around one of two directions of the axis **n** of easiest magnetization.

If the frequency of the rf field is chosen to be ω_g or ω_e, then it is easy to show that, to within terms quadratic in h_0, a system of quasilevels (see Fig. 7a) determined by equations of the type (32) and (35) appears in complete analogy to the results obtained in the preceding section. In the case when $\omega_{\rm rf}=\omega_e$,

$$\omega_{\tilde{m}}=\frac{1}{2}h_0\omega_g\tilde{m}\sin\theta,\quad \omega_{\tilde{M}}=0,\qquad \omega_m=\omega_g m,\quad \omega_M=\omega_e M. \quad (37)$$

Here m is the projection of the nuclear spin on the **n** axis, and $\tilde{m}$ is the projection of the nuclear spin on an axis perpendicular to **n** and lying in the **n**–**H** plane. Similarly, for $\omega_{\rm rf}=\omega_e$

$$\omega_{\tilde{m}}=0,\quad \omega_{\tilde{M}}=\frac{1}{2}h_0\omega_e\tilde{M}\sin\theta,\qquad \omega_m=\omega_g m,\quad \omega_M=\omega_e M. \quad (38)$$

The absorption spectrum is then described by Eq. (36) obtained in the preceding section. However, it should be kept in mind that in the weak-field case under study Eq. (36) is approximate, since it neglects corrections that are quadratic in the amplitude of the rf field. Figures 7b–d show the Mössbauer spectra at rf frequencies near resonance values. As one can see from these figures, the splitting of the hyperfine lines

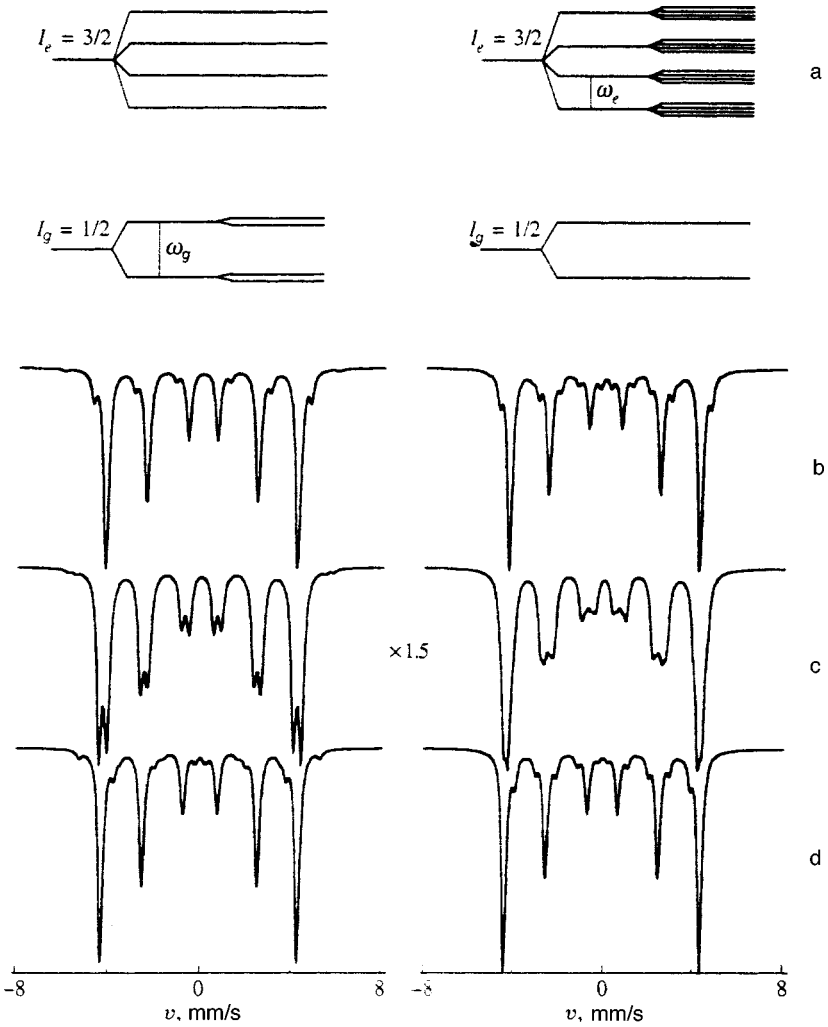

FIG. 7. a—Scheme of the splitting of the energy levels of a nucleus in excited and ground states in a weak rf field ($h_0=0.25$) at resonance frequencies $\omega_{\rm rf}=\omega_g=36.8$ MHz (left side) and $\omega_{\rm rf}=\omega_e=21.1$ MHz (right side). The corresponding Mössbauer spectra for an ensemble of randomly oriented SW particles in a rf field with frequency close to the resonance values ($\Delta\omega=5$ MHz): $\omega_{\rm rf}=\omega_{g,e}+\Delta\omega$, $\omega_{g,e}$, $\omega_{g,e}-\Delta\omega$ (b–d). The scale of the spectra in (b) is magnified by a factor of 1.5.

disappears when the rf frequency increases or decreases relative to the resonance frequency. We also note that traces of this splitting remain quite far from the resonance frequencies, i.e., for $\Delta\omega_{\rm rf}$ much greater than the natural width Γ_0. Such line splittings should be observed not only for $\theta=90°$ but also for smaller angles, for which the magnitude of the splitting decreases as $\sin\theta$. In consequence, these splittings remain in the total spectrum summed over all possible orientations of the axis of easiest magnetization (see Fig. 5).

7. CONCLUSIONS

The theory developed above makes it possible to calculate the Mössbauer absorption spectra in the case when the hyperfine field at a nucleus changes direction arbitrarily. Specific calculations were performed and the transformation of the hyperfine spectra as a function of the frequencies and amplitudes of the rf field were analyzed for the Stoner–Wohlfarth model in strict correspondence to the original postulates of this model.[16] A series of qualitatively new effects was observed in the hyperfine spectra as predicted by the present model. These effects are completely absent if magnetization reversal occurs as a sharp reversal of the direction of the axis of easiest magnetization.[15] To understand the characteristics of the observed effects it was found that it is important to study the case of a circularly polarized hyperfine field, for which an analytical solution was obtained. The hyperfine spectrum in this case consists of Lorentzian lines of natural width, the number of lines, however, being much greater than in a static hyperfine field.

It should be kept in mind that the Stoner–Wohlfarth model may be inadequate for describing magnetization reversal in real situations, first and foremost because this model neglects the interparticle interaction and also because the relaxation process is described very approximately. These questions require special analysis.

*) E-mail: chuev@labdif.crystal.msk.su

[1] L. Pfeiffer, J. Appl. Phys. **42**, 1725 (1971).
[2] L. Pfeiffer in *Mössbauer Effect Methodology*, edited by I. J. Gruverman, Plenum Press, N.Y., 1972, Vol. 7, p. 263.
[3] G. Albanese and G. Asti, Nuovo Cimento B **6**, 153 (1971).
[4] M. Kopcewicz, J. de Phys. Coll. C **6**, C6-107 (1976).
[5] M. Kopcewicz, Phys. Status Solidi A **46**, 675 (1978).
[6] M. Kopcewicz and A. Kotlicki, J. Phys. Chem. Solids **41**, 631 (1980).
[7] M. Kopcewicz, U. Gonser, and H.-G. Wagner, Nucl. Instrum. Methods **199**, 163 (1982).
[8] M. Kopcewicz, Struct. Chem. **2**, 313 (1991).
[9] T. Graf, M. Kopcewicz, and J. Hesse, Nanostruct. Mater. **6**, 937 (1995).
[10] M. Kopcewicz, A. Grabias, and P. Nowicki, Nanostruct. Mater. **6**, 957 (1995).

[11] T. Graf, M. Kopcewicz, and J. Hesse, J. Phys.: Condens. Matter **8**, 3897 (1996).
[12] S. Olariu, I. Popesku, and C. B. Collins, Phys. Rev. C **23**, 50, 1007 (1981).
[13] S. R. Julian and J. M. Daniels, Phys. Rev. B **38**, 4394 (1988).
[14] A. Yu. Dzyublik, Phys. Status Solidi B **194**, 699 (1996).
[15] A. M. Afanas'ev, M. A. Chuev, and J. Hesse, Phys. Rev. B **56**, 5489 (1997).
[16] E. C. Stoner and E. P. Wohlfarth, Philos. Trans. R. Soc. London, Ser. A **240**, 599 (1948).
[17] W. Heitler, *Quantum Theory of Radiation*, Clarendon Press, Oxford, 1954.
[18] R. Zwanzig, Physica **30**, 1109 (1964).

Translated by M. E. Alferieff

Critical behavior and mechanism of strain correlations under conditions of unstable plastic flow

M. A. Lebyodkin*) and L. R. Dunin-Barkowskiĭ†)

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

(Submitted 27 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 1816–1829 (May 1998)

The complicated time dependence of discontinuous deformation and statistical properties of stress drops associated with the Portevin–Le Châtelier effect in single crystals and polycrystalline samples of the Al–Mg alloy have been studied experimentally. We have determined conditions under which the stress drops have no characteristic scale, but display a power-law distribution. Such statistical features of critical behavior have been observed in systems with various types of plastic strain dynamics, namely, in propagation of deformation bands and avalanche slips, which demonstrate some properties of models of self-organized criticality. The spatial correlation in processes of collective dislocation motion is determined by inhomogeneous elastic stresses in the dislocation ensemble. A computer model taking into account local nonlinearity of crystal resistance to plastic flow under conditions of the Portevin–Le Châtelier effect and correlations between strains in different elements provides an adequate description of both the statistical properties of the effect and the strain distribution in space and time. © *1998 American Institute of Physics.* [S1063-7761(98)01805-8]

1. INTRODUCTION

The creation of the theory of dynamic systems[1] has resulted in considerable changes in approaches to various phenomena studied in different fields of science, ranging from physics to biology. An ensemble of dislocations is an example of a dynamic dissipative system with interacting components, whose evolution displays some features of self-organization in time and space. Several review articles have been devoted to collective behavior of dislocations.[2,3] The phenomenon of plastic flow instability, i.e., discontinuous deformation due to abrupt coherent glide of large dislocation groups, has been widely known. This process can be caused by various microscopic mechanisms. The behavior of strain in time and space, however, often demonstrates some universal features, independent of the nature of instability.[4] On the other hand, a large variety of stress–strain curves, ranging from fairly regular to those typical of nonstationary random processes, can be observed for the same mechanism, depending on deformation conditions.[5–7] Different types of macroscopic behavior correspond to radically different dynamic properties of moving dislocations. This variety is no less interesting than their common properties. Attention has been attracted to studies of plastic deformation as a model of cooperative behavior of dissipative systems by the possibility of changing the dynamic state of a dislocation ensemble by varying parameters of the experiment, microstructure, and sample dimensions.

The behavior of discontinuous deformation of diluted alloys due to the Portevin–Le Châtelier effect[8] is probably the most versatile. The microscopic mechanism of this effect has been studied in detail.[9] The cause of instability in the plastic deformation in this case is a section with negative sensitivity on the curves of stress σ versus strain rate $\dot{\epsilon}$ caused by diffusion of impurities to dislocations arrested on obstacles (in other words, dynamic strain ageing), i.e., a lower stress σ is needed at a higher plastic strain rate $\dot{\epsilon}$. The reason for such an effect is quite clear: the lower the plastic strain rate, the longer the time during which a dislocation is pinned on obstacles, therefore the more impurity atoms drift to the dislocation and increase the potential barrier to the dislocation unpinning. The competition between the normal, positive strain rate sensitivity owing to thermally activated processes and the negative component due to impurities gives rise to an N-shaped stress characteristic of σ versus strain rate $\dot{\epsilon}$ (Fig. 1). In the ideal case of homogeneous strain, the negative resistance of a crystal to plastic strain, like the negative differential electric resistance, should lead to periodic oscillations in the plastic strain rate similar to relaxation oscillations in electric circuits.[10] When the testing machine is designed to produce a constant strain rate, oscillations in $\dot{\epsilon}$ translate into oscillations in σ owing to the elastic reaction of the testing machine. The ideal, periodic shape of oscillations is never observed because the force coupling dislocations to one another is finite, thus the latter parameter is an important factor in the plastic slip dynamics and irregularities in stress–strain curves.

There is a profound similarity between the Portevin–Le Châtelier effect and instabilities in dry friction, which is based on the paradigm of self-organized criticality in dynamic systems with large numbers of degrees of freedom.[11,12] Theoretical models of dry friction have been used in attempts to account for the statistics of earthquakes, based on the assumption of an anomalous negative[12] dependence of the friction coefficient on the relative slip velocity. Instabili-

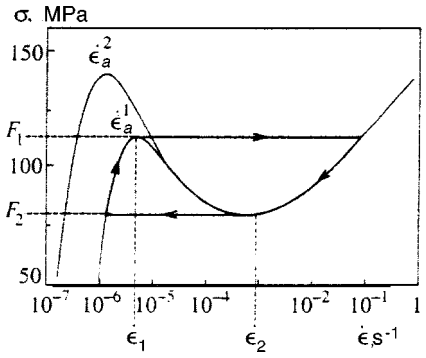

FIG. 1. Flow stress as a function of plastic strain rate, $F(\dot\epsilon)$, for two values of the applied strain rate $\dot\epsilon_a$ ($\dot\epsilon_a^1 > \dot\epsilon_a^2$).

ties of this type lead to avalanche processes whose magnitudes have no characteristic scale, but are described by power-law distributions, and in this sense, the system is in a critical state. The possibility of self-organized critical states due to various mechanisms of low-temperature stress jumps was suggested by Bobrov et al.,[4] and this conclusion was based on the observation of common features in statistics of electric signals caused by the twinning and catastrophic slip. The first studies of statistical properties of the Portevin–Le Châtelier effect revealed that it was characterized by a power-law distribution in a certain range of deformation parameters, but this was not the end of the story.[5–7] An alternative approach to the dynamic model of the effect is based on the possibility of chaotic behavior under conditions of the Portevin–Le Châtelier instability, which was predicted theoretically[13] and later confirmed in experiments.[14] Unlike the case of self-organized criticality, several degrees of freedom, which, obviously, correspond to collective degrees of freedom in a dislocation ensemble, are sufficient for the occurrence of deterministic chaos. These two concepts do not rule out various types of correlated dislocation behavior, depending on the deformation conditions.

This paper describes a statistical approach to quantitative analysis of stress–strain curves under conditions of unstable plastic flow. Effects of strain rate, temperature, microstructure, and the size of the Al–Mg alloy sample on the shape of deformation curves and statistics of stress drops have been studied. The paper presents the data concerning the microscopic mechanism of spatial correlations in deformation processes and conditions under which dynamic and statistical properties of the Portevin–Le Châtelier effect have a critical character. We have studied the nonlocal dynamics of a computer model which represents a strained crystal as a one-dimensional chain of elements whose plastic flow is described by a local nonlinear material equation.

2. EXPERIMENTAL TECHNIQUES

Polycrystalline samples of the Al–3 at.% Mg alloy and single crystals of the Al–4.5 at.% Mg alloy were cut in a standard shape for tensile experiments (symmetrical paddles) and strained on a rigid testing machine (the machine-sample stiffness was $C \approx 10^7$ N/m) at constant grip velocities corresponding to sample strain rate $\dot\epsilon_a = 2\cdot 10^{-6} - 2\cdot 10^{-3}$ s^{-1} at temperatures of 20, 80, and 120° C. The length l, width d, and thickness w of the studied section of polycrystalline samples were in the ranges $l = 18–36$ mm, $d = 1–6$ mm, and $w = 0.5–2$ mm. Dimensions of single crystal samples were $l = 7–28$ mm, $d = 1–5.5$ mm, and $w = 0.7–1.5$ mm. The alignment of the tensile axis in single crystals close to the $\langle 111\rangle$ or $\langle 100\rangle$ directions corresponded to the case of multiple slip. In order to study the impact of the microstructural state, polycrystalline samples shaped by cold rolling to a plastic strain value of $\epsilon = 0.5$ were recrystallized by anneals at different temperatures in the range of 360° to 460°C. The average grain size varied between 50 and 500 μm, depending on the anneal temperature.

Since the microstructural state of crystals changes in the process of straining, which manifests itself as a change in the strain hardening rate, an ideal statistical procedure should accumulate measurements from a narrow range around a fixed strain ϵ for a statistically large set of samples with equal hardening coefficient, and it would be equivalent to averaging over a statistical ensemble. This was impossible to do in real experiments. For this reason, when the average stress drop magnitude $\Delta\sigma$ varied systematically as a function of ϵ, which it did linearly in our experiments to a fair accuracy, the stress drop amplitude was normalized to the linear function $f(\epsilon)$ obtained using a linear least-squares fit through points of $\Delta\sigma$ versus ϵ. Then the distributions of the normalized stress drop magnitudes, $s = \Delta\sigma / f(\epsilon)$, were plotted.

The impact of microstructure was manifested in variations in distribution shapes from sample to sample. Therefore statistical data were accumulated in each experiment for several strain rates $\dot\epsilon_a$ at a fixed temperature or for different temperatures at a fixed $\dot\epsilon_a$. The robustness of the distribution shape during an experiment was tested by recording repeatedly a section of the stress–strain curve at a parameter value selected as a reference. It turned out that the statistics character varied qualitatively with the deformation conditions similarly in all samples. The quantity of statistical data for each set of parameters was 100 to 300 stress drops.

3. EXPERIMENTAL RESULTS

Studying statistical distributions of stress drop magnitudes allows one to estimate deviations of a real stress–strain curve from the ideal case of regular relaxation oscillations with a constant amplitude. The underlying tendency in statistics of stress drops changing with experimental conditions is a gradual transition from bell-shaped, almost normal distributions to asymmetric, monotonically dropping curves. At the same time, the shapes of the stress–strain curves change radically, which corresponds to changes in the pattern of localized slips in a strained sample (see also Ref. 15). This section presents a description of the impact of various experimental parameters on statistical stress drop distributions and spatial distribution of strain in the process of experiment.

The impact of the microstructural crystal state is seen most clearly in comparing the initial deformation stage, which is characterized by a nonequilibrium dislocation microstructure, with the later stage corresponding to an almost stable microstructure, and also in comparison between data

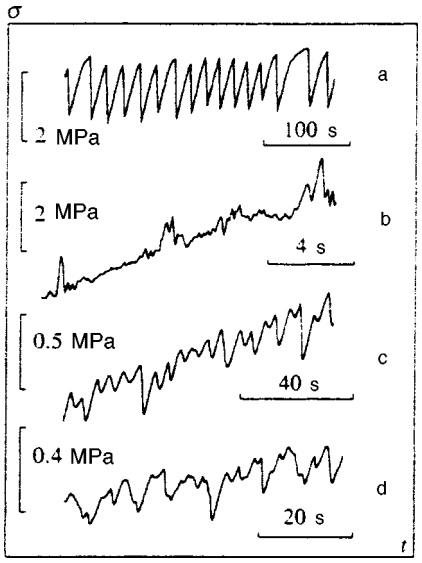

FIG. 2. Sections of stress–strain curves at room temperature on the stage of quasi-stable jump-like deformation of a polycrystal annealed at 400° C at strain rates (a) $\dot{\epsilon}_a = 1.3\times10^{-5}$ s^{-1} and (b) 5.3×10^{-4} s^{-1}, and of single crystals of width (c) $d=4$ mm and (d) 1.2 mm at $\dot{\epsilon}_a = 1.3\times10^{-5}$ s^{-1}.

from single crystals and polycrystals annealed at different temperatures. At strains of $\epsilon < 3$–5% for single crystals and 6–8% for polycrystals, the strain hardening is rapid, and stress drops are characterized by large spreads in the distributions of their amplitudes and times at which they occur. On the later stage, when the average hardening coefficient is close to zero (in this sense, this is a quasisteady stage), the stress drop characteristics are more regular. Figure 2 contains curves a and c, which are sections of stress–strain curves in this stage of experiments with poly- and single crystals at a strain rate near the lower edge of the studied interval of $\dot{\epsilon}_a$. Such curve shapes correspond to localization of strain in slip bands. As follows from optical measurements of sample surfaces, each stress drop correlates with formation of a localized slip band with a typical width of about one millimeter. The regularity of stress–strain curves in the stage of repeated stress drops is seen in the presence of a characteristic amplitude of stress drops, namely the peak position on the distribution histogram (Fig. 3a and 3b). The initial deformation stage is characterized by a wider peak in the distribution (polycrystals) or monotonically dropping distributions (single crystals). Thus, the decrease in the material plasticity owing to strain hardening leads to narrower distributions of stress drop magnitudes. The smaller the grain size in annealed polycrystals, the less notable is this effect. It is almost unobservable in nonannealed samples.

The analysis of the effect of sample pretreatment after cold rolling of polycrystals leads to a similar conclusion concerning the effect of plasticity. The shapes of stress–strain curves (for example, curves a and c in Fig. 2) and distributions (Fig. 3a and 3b) for different samples in the same stage of irregular deformation have been compared. Differences between distributions for poly- and single crystals in the initial stage were mentioned above. Under conditions of stabilized plastic flow, the narrowest distributions are observed in nonannealed polycrystals. In tests of annealed polycrystals or single crystals, the probabilities of jumps of smaller magnitudes are notably higher, and the distribution center of gravity shifts toward smaller stress drop amplitudes.

The effects of strain rate, temperature, and sample dimensions were studied in the stage of stabilized discontinuous deformation. The character of stress instability changes with $\dot{\epsilon}_a$ (Fig. 2a and 2b), which corresponds to a transition from dislocation slip localization in immobile deformation bands (slip bands) to formation of bands propagating through a crystal. This effect has been widely known in polycrystals, which have stress–strain curves and spatial correlations in processes of dislocation slip of clearly distinguishable types.[7,15] Stress–strain curves of single crystals are less regular, and no accurate classification can be applied to them. In this connection, the data on single crystals are sparse in the literature. It follows from experimental data that the underlying tendency in the effect of strain rate on stress–strain curves and $\Delta\sigma$ distributions is the same in single and polycrystals. For this reason, the statistics of stress drops presents convenient numerical characteristics of changes due to variations in the deformation conditions.

The statistical analysis indicates that distributions of stress drop amplitudes gradually become broader with increasing $\dot{\epsilon}_a$ because drops of various amplitudes turn up. Beyond some strain rate, which depends on the temperature and sample structure, monotonically decreasing distributions are observed (Fig. 4a). In this case, stress drops are related to the nucleation of deformation bands, and to fluctuations in their widths and velocities. Comparison of data derived from different samples reinforces the above conclusion about the effect of microstructure based on measurements at low strain rates: the higher the material plasticity, the smaller the strain rate $\dot{\epsilon}_a$ beyond which the distribution is described by a monotonically decreasing function. An important feature of monotonic distributions is that over wide ranges of $\dot{\epsilon}_a$ they

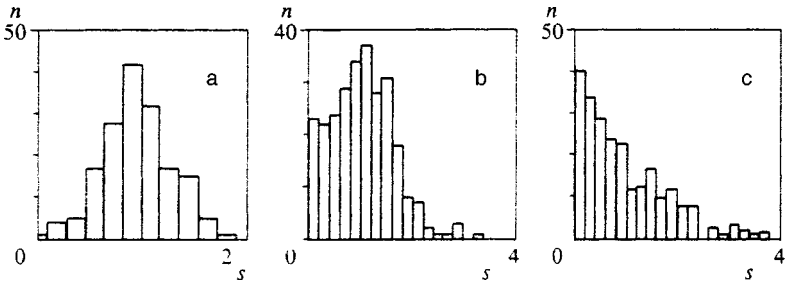

FIG. 3. Distribution histograms of normalized stress jump amplitudes at $\dot{\epsilon}_a = 1.3\cdot10^{-5}$ s^{-1} for (a) a polycrystal annealed at 400° C, and single crystals of width (b) $d=4$ mm and (c) 1.2 mm.

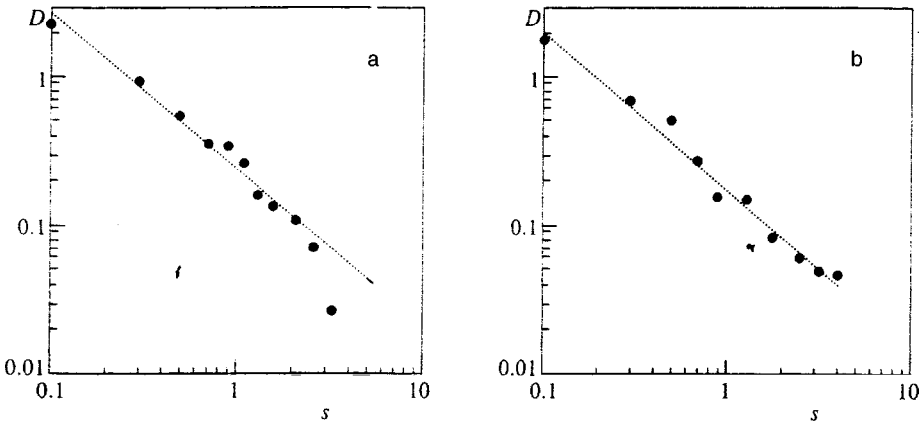

FIG. 4. Distribution functions $D(s)$ of normalized stress drop amplitudes for (a) a polycrystal tested at $T=20°$ C and $\dot{\epsilon}_a=8.2\times10^{-4}$ s^{-1} and (b) a single crystal at $T=120°$ C and $\dot{\epsilon}_a=1.3\times10^{-5}$ s^{-1}. The slopes of straight lines correspond to the following exponents α in functions $D(s)\sim s^{\alpha}$: (a) $\alpha\approx-1.08\pm0.06$ and (b) $\alpha\approx-1.06\pm0.08$.

can be approximated with fairly good accuracy by power functions $D(s)\sim s^{\alpha}$, where α varies from -1 to -1.5. One example of $D(s)$, the distribution of normalized stress drop amplitudes approximated by a power function, is given in Fig. 4a. The absence of a characteristic stress drop scale from these processes allows us to speak about a critical state of the dislocation system.

Another feature providing evidence in favor of criticality was observed in deformation of single crystals when the temperature was increased to 120° C and the strain rate corresponded (see the caption to Fig. 3) to bell-shaped distributions of $\Delta\sigma$ at room temperature. Under these conditions, stress–strain curves of single crystals have a distinctive form characteristic of slip localization in immobile deformation bands, although the spread in $\Delta\sigma$ is quite considerable. Nonetheless, the resulting distributions are described by power functions $D(s)$ (Fig. 4b) with the exponent α in the same range as in propagation of deformation bands at high $\dot{\epsilon}_a$.[1)] A similar behavior was observed in recent experiments on direct detection of localization and propagation of deformation bands at room temperature in polycrystals with an average grain size of up to several millimeters. However, unlike the case of propagation, power-law distributions associated with formation of localized bands were detected only in narrow ranges of T and $\dot{\epsilon}_a$. A more detailed assessment of the impact of temperature on the Portevin–Le Châtelier effect is given elsewhere.[16]

An investigation of the effect of sample length was interesting in the context of the self-organized criticality model,[11,12] characterized by dimensional scaling of the statistics. Notwithstanding the predictions derived from these models, the distributions of stress drop amplitudes under conditions of the Portevin–Le Châtelier effect proved to be insensitive to changes in sample lengths in the studied interval. At the same time, a reduction in the single crystal width had the same impact on stress–strain curves and distributions as an increase in the sample plasticity (Figs. 2d and 3c). No such effect of sample dimensions was detected in polycrystals.

The impact of deformation and transverse sample dimensions on the statistics of stress drops allows us to make some statements about the feasible mechanism relating slip processes in conditions of deformation instability. In fact, the existence of a most probable stress drop amplitude, at which distributions of $\Delta\sigma$ in massive samples at high strain are centered, indicates that correlations in the dislocation system are stronger than in plastic or thin crystals, which display stress drops of various amplitudes. This can be explained assuming that strain can propagate to neighboring slip planes owing to elastic stresses due to inhomogeneous plastic strain. Plastic relaxation of elastic stress fields should reduce the effective force coupling neighboring elements. Therefore, the impact of plastic strain can be due to less favorable conditions for plastic relaxation owing to production of obstacles to dislocation motion, and, in contrast, the reduction in the transverse sample dimension is favorable for relaxation due to the dislocation escape through the sample surface. No size effect has been detected in polycrystals because grain boundaries are efficient barriers to the dislocation motion. Note that the effects of temperature and initial strain rate are consistent with our conclusions, but the analysis of these effect requires that changes in the shape of the stress vs strain rate curve should be also taken into account (see below).

The nature of elastic stresses in inhomogeneous plastic slip was first explained by Eshelby.[17] The point is that the mismatch between plastic strains in neighboring regions of a crystal should be compensated for by elastic strains so that the condition of full strain continuity should be satisfied. Other mechanisms connecting neighboring elements of a strained material were also discussed,[15,18] such as transfer of dislocations to neighboring planes owing to double cross-slip, nonlocal hardening in a deformation band, and, finally, breaking of the stress unidimensionality due to changes in the sample shape in the process of testing. The qualitative predictions derived from these models[15,18] are in conflict with observations of the effects of sample dimension and strain on the statistics of stress drops. Estimates of characteristic coupling forces given in Refs. 15 and 18 also indicate a dominant role of elastic interaction. This allows us to develop a simple model of inhomogeneous serrated deformation.

4. NUMERICAL SIMULATION

Local dynamic properties of the Portevin–Le Châtelier effect are approximately described in the context of disloca-

tion motion in a crystal cross section using a nonlinear material equation relating the stress to plastic strain ϵ and strain rate $\dot{\epsilon}$[19]:

$$\sigma = h\epsilon + F(\dot{\epsilon}). \tag{1}$$

The hardening coefficient h characterizing the increase in the nonthermal component of σ is assumed to be constant for simplicity. Dislocation models (such as that in Ref. 9) predict an N-shaped curve of $F(\dot{\epsilon})$, which is a result of interaction between mobile dislocations and local obstacles (Fig. 1). The phase portrait of the system plotted in $(\sigma, \dot{\epsilon})$ coordinates, shown schematically in Fig. 1, is a cyclic orbit including two horizontal jumps (relaxation oscillations[10]). In (σ,ϵ) coordinates this orbit corresponds to a periodic discontinuous stress–strain curve. In order to take account of the inhomogeneity of plastic flow, the sample was modeled by a one-dimensional chain of N blocks, which are small elements of the sample where deformation can be considered to be quasihomogeneous. Such regions are probably bundles of slip lines, which are formed quasisimultaneously with respect to the time when deformation bands are formed. The one-dimensional model was selected on the base of the experimental fact that generation of bands is much faster than their axial propagation. Given the experimental data indicating that the nature of correlation among strained elements of the sample is elastic, we can simulate coupling between neighboring elements by springs of strength K. If the strain in a specific block is inconsistent with that of its neighbors, a ''backstress'' is generated which tends to equalize plastic strain rates. Now the modified material equation for the ith block reads as follows:

$$\sigma = h\epsilon_i + F(\dot{\epsilon}_i) + K((\epsilon_i - \epsilon_{i-1}) + (\epsilon_i - \epsilon_{i+1})). \tag{2}$$

In the continuum approximation, this modification is equivalent to the additional term $C\delta_x^2\epsilon$ (Ref. 20) on the right-hand side of Eq. (1). Here C is the coupling constant (first derivatives are not included because of the invariance under sample rotations). The constants K and C are related by the formula $K = Ca^2$, where a is a characteristic length which has the sense of interaction range.

This equation should be supplemented with a relationship characterizing deformation conditions. The strain rate $\dot{\epsilon}_a$, controlled by the testing machine, has two components, namely the elastic strain rate in the machine–sample system, described by Hooke's law, and the plastic strain rate in the sample, determined in the case of inhomogeneous strain by averaging over the sample:

$$\dot{\epsilon}_a = \frac{\dot{\sigma}}{M} + \frac{1}{N}\sum \dot{\epsilon}_i, \tag{3}$$

where M is the effective system stiffness (the sample elastic modulus in the ideal case of a machine with infinite rigidity).

The shape of the $F(\dot{\epsilon})$ curve was calculated using the microscopic model.[9,21] The amplitude of the anomalous section of the curve decreases with $\dot{\epsilon}_a$, and the peak position $\dot{\epsilon}_1$ shifts towards higher $\dot{\epsilon}$ (Fig. 1). Similarly, $\dot{\epsilon}_1$ increases with temperature,[7,22] although it is difficult to calculate accurately because the thermal activation parameters are poorly known. Details of the numerical solution of the equation system are given elsewhere.[7] The model parameters were varied within the intervals $K = (0.03–1.0)M$, $\dot{\epsilon}_a = (0.01–0.6)\dot{\epsilon}_2$, $h = (0–0.01)M$, and $N = 25–600$. Inhomogeneity was introduced to the system through random initial values of $\dot{\epsilon}_i$ fluctuating within an interval $0.05\dot{\epsilon}_1$. Results for each random set of parameters were accumulated after achieving statistically stable states of the discontinuous deformation without additional noise.

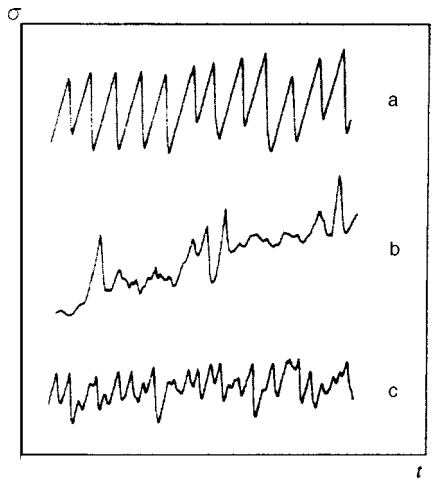

FIG. 5. Examples of calculated stress–strain curves at (a) $K = 0.5M$, $\dot{\epsilon}_a = 0.05\dot{\epsilon}_2$, (b) enhanced strain rate $\dot{\epsilon}_a = 0.3\dot{\epsilon}_2$, and (c) reduced coupling constant $K = 0.15M$.

The model is essentially different from the earthquake model describing motion of a chain of stiff blocks.[12] First, our model considers plastic strain in a system of connected blocks, each of which is characterized by its values of ϵ_i and $\dot{\epsilon}_i$, which are solutions of the plastic flow equations (2) and (3). An important consequence is that one can use the nearest-neighbor approximation, since the plastic strain rate is an exponential function of stress. Second, the frictional force in the earthquake model is a monotonically decreasing function of velocity. If an N-shaped curve of the flow resistance function based on the microscopic theory is used, there is a characteristic time scale related to the period of relaxation oscillations in a homogeneous sample.

5. RESULTS OF COMPUTER SIMULATION

A solution to the system of equations (2) and (3) was sought in the form of functions $\sigma(t)$ and $\dot{\epsilon}_i(t)$. Curves of $\sigma(t)$ were analyzed similarly to experimental data. The function $\dot{\epsilon}_i(t)$ characterizes dynamic properties of the model, namely the evolution of the strain rate distribution in space. It was found that the numerical solution is determined qualitatively by parameters K and $\dot{\epsilon}_a$, whereas variations in h only slightly shift the intervals of K and $\dot{\epsilon}_a$ in which a feature occurs. Below we give results of a computer simulation at $h = 0.01M$, which is a value typical of the alloys under investigation.

Figure 5 shows examples of calculated stress–strain

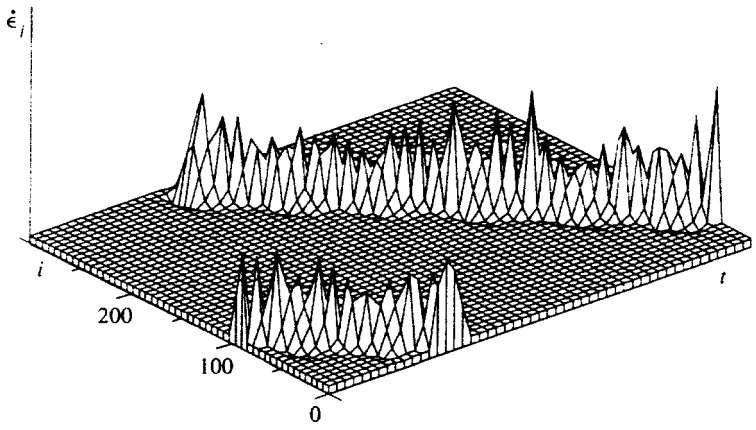

FIG. 6. Plastic flow rate of the ith chain element as a function of its position i and time t; $K=0.5M$, $\dot{\epsilon}_a=0.3\dot{\epsilon}_2$. The surface shows fluctuations in the rate and width of the excitation (''deformation band'') propagating along the chain.

curves for three cases. In the interval $K\sim(0.5–1)M$ the qualitative features of the model are independent of K. At deformation rates below $\dot{\epsilon}_a\sim 0.1\dot{\epsilon}_2$ the solutions yield regular curves of $\sigma(t)$ (Fig. 5a), which are characterized by bell-shaped distributions of stress drop amplitudes. At lower K, jumps of smaller amplitudes occur alongside large drops, so the width $\Delta\sigma$ of peaks in the distributions increases. Each stress drop is due to an abrupt increase in $\dot{\epsilon}$ caused by a sudden displacement $\dot{\epsilon}_i$ of a group of adjacent blocks (ten to thirty blocks at high stress drop amplitudes) to the right-hand rising section of the $F(\dot{\epsilon})$ curve. This is the process by which a localized deformation band is created. At higher $\dot{\epsilon}_a$, curves (Fig. 5b) with features similar to those of polycrystal curves (Fig. 2b) are observed. The system dynamics can be depicted graphically by a surface in ''time — block number i — plastic flow rate in the block'' coordinates plotted in Fig. 6. One can see that, in these conditions, the ''deformation bands'' propagate through the crystal, i.e., regions with high plastic flow rates, $\dot{\epsilon}_i>\dot{\epsilon}_2$. The smaller the K, the lower the $\dot{\epsilon}_a$ at which localization is replaced by propagation of deformation bands. The nucleation of bands, fluctuations in their velocities and width cause variations in stress σ. Over a wide range of strain rates, jump amplitudes $\Delta\sigma$ are distributed in accordance with a power law with exponent α from the interval determined in experiments. One example of the jump amplitude distribution obtained in computer simulations is shown in Fig. 7a.

Propagation of deformation bands at high $\dot{\epsilon}_a$ in the example with K close to M is the feature common to the model dynamics at all K. With a view to testing qualitative conclusions about the nature of spatial coupling, it is interesting to analyze changes due to a decrease in K under conditions of localized strain. In the interval of K about $0.3M$, distributions of stress drop amplitudes correspond to those observed in single crystals. When K decreases further, bell-shaped distributions are replaced by monotonically decreasing functions. In this case features typical of stress–strain curves of single crystals at initial stages and of thin crystal deformation are reproduced (Fig. 5c). In narrow ranges of K and $\dot{\epsilon}_a$, power-law distributions of stress drop amplitudes are observed, i.e., there is no characteristic scale of stress drop amplitude. Figure 8 illustrates dynamics of the model in these conditions, which displays avalanches in the plastic flow. A curve of $D(s)$ for this case if given in Fig. 7b, which is described by a power function because dimensions of avalanches are distributed throughout the range of allowed dimensions in the system. Such avalanches are characteristic of self-organized criticality. However, unlike previously discussed models,[11,12] fine tuning of system parameters is needed to simulate such a deformation regime in our model.

In the cases of both localized strain and propagation of deformation bands, the power-law distributions are weakly susceptible to the number of blocks in the chain. The distributions deviate from power functions only for $N<75$, and

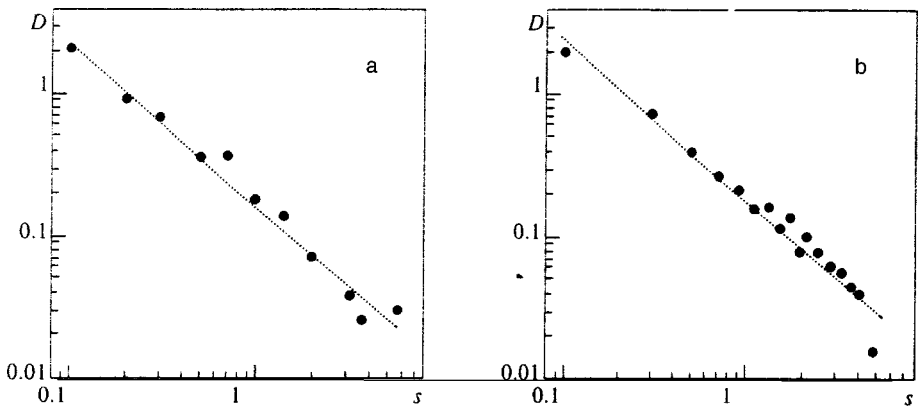

FIG. 7. Distribution functions $D(s)$ calculated at the same parameters as (a) in Fig. 6 and (b) in Fig. 8: (a) $\alpha\approx-1.14\pm0.07$; (b) -1.12 ± 0.06.

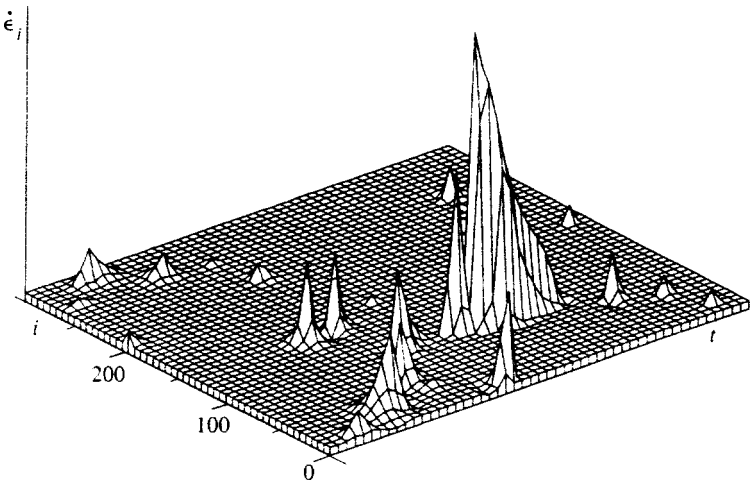

FIG. 8. Profile of the plastic flow rate in a chain of blocks for the case of avalanche slip processes: $K=0.12M$, $\dot{\epsilon}_a=0.05\dot{\epsilon}_2$.

the exponent α remains constant. This result is consistent with the absence of a notable sample size effect in experimental data.

Our results indicate that the model provides an adequate description of spatial patterns of strain localizations and statistical distributions of stress drop amplitudes under conditions of the Portevin–Le Châtelier effect. The behavior of the least plastic samples, namely cold-rolled polycrystals, corresponds to the coupling constant K close to the elastic modulus. At lower K the model reproduces changes observed in experiments with annealed polycrystals, single crystals of smaller cross sections, and at higher testing temperatures. This facts favor our conclusion about the elastic nature of correlations in processes of discontinuous deformation derived from the qualitative analysis of experimental data.

It is clear that the temperature effect can be easily interpreted in terms of a lower correlation parameter K. Let us analyze in greater detail the role of changes in the curve of $F(\dot{\epsilon})$ due to variations in the strain rate and temperature. It follows from our computer simulations that at given K the dynamics of deformation bands and shapes of distributions of $\Delta\sigma$ depend on the strain rate controlled by the testing machine as compared to positions of peaks on the curve of $F(\dot{\epsilon})$. The onset of band propagation corresponds to a strain rate close to the upper edge of the instability domain, corresponding to the minimum of the function. The peak position $\dot{\epsilon}_1$ shifts towards the minimum $\dot{\epsilon}_2$ as $\dot{\epsilon}_a$ or temperature increases (Fig. 1). But in the former case this shift is slower than the increase in $\dot{\epsilon}_a$, which leads to the regime of band propagation at sufficiently high $\dot{\epsilon}_a$. When the temperature increases, $\dot{\epsilon}_a$ is fixed, so its position becomes closer to $\dot{\epsilon}_1$, which should result in strain localization and peaked stress drop distributions. Therefore, it is feasible that the disappearance of the characteristic jump amplitude with temperature increase is due to the decrease in the coupling force caused by plastic relaxation of elastic fields. Since high-temperature features of the Portevin–Le Châtelier effect are reproduced by the model at the lowest K, this is another piece of evidence in favor of the elastic coupling hypothesis.

The good agreement between the model and experimental results allows us to suggest a mechanism that determines dynamic properties of the Portevin–Le Châtelier effect. The model dynamics is controlled by the competition between the characteristic time of motion on sections of stable plastic flow on the phase trajectory (Fig. 1) and the time in which the plastic flow rates in the sample are equalized. When the strain rate is low and the coupling force high, different elements of the sample have close values of $\dot{\epsilon}_i$, therefore instability in one element leads to an avalanche-like formation of a deformation band, which terminates when the deforming stress drops. In the opposite case, an inhomogeneous strain gradient is maintained in the sample, for this reason, periodically repeated propagation of the plastic slip zone through the sample takes place.

6. CONCLUSIONS

Thus, we have obtained in this work new data providing evidence in favor of the elastic nature of correlations among deformation processes in the Portevin–Le Châtelier effect and revealing two regimes in which the dislocation system is self-organized to a state with a power-law distribution of elementary process energies. These regimes are characterized by different dynamic features of the dislocation system, namely, generation of localized or propagating deformation bands, and probably these two processes are driven by two different physical mechanisms. The statistical analysis has not clarified the nature of criticality, but it has allowed us to detect the critical behavior of plastic flow. In the case of immobile bands, the dynamics and statistics of the Portevin–Le Châtelier effect are similar to those of a dissipative system with multiple degrees of freedom and characterized by a tendency to self-organized criticality. On the other hand, such a state has been detected in our systems only in narrow ranges of deformation parameters, unlike the case of self-organized criticality. This difference is probably caused by the existence of a characteristic time for processes in the Portevin–Le Châtelier effect, namely the period of relaxation oscillations, which leads to the presence of different statistical regimes, including those with a characteristic time scale.

The important features of the computer model developed in this work are the negative derivative of stress with respect to the strain rate, which is responsible for the instability of plastic flow, and the mesoscopic spatial scale deriving from the assumption about elastic coupling among incoherently strained material elements. The latter feature results in complex spatial distributions of strain in the material. The good agreement between computer simulations and experimental data allows us to hypothesize that in spite of the complexity of plastic flow processes these two features determine the principal dynamic properties of the dislocation system under conditions of the Portevin–Le Châtelier effect. The changes due to variations in deformation parameters include avalanche-like plastic flow in different parts of the crystal and soliton-like propagation of regions with localized strain. Although our model is fairly simple, it may be used in modeling real systems with instabilities of this type. In this connection, note a recent publication[23] in which an N-shaped curve of resistance versus velocity was suggested for models of various excited media, such as systems with dry friction, electric power lines, and optical waveguides.[23]

We are thankful to V. Ya. Kravchenko, Y. Bréchet, Y. Estrin, and L. P. Kubin for interest in the work and helpful discussions, and to A. P. Ivanov for technical help. The work was supported by the Russian Fund for Fundamental Research (Project 96-04-17482) and ISF (Grant NKV300).

*)E-mail: lebedkin@issp.ac.ru
†)E-mail: dbarkov@issp.ac.ru

1)Similar behavior was observed in recent experiments on direct detection of localization and propagation of deformation bands at room temperature in polycrystals with an average grain size of up to several millimeters.

[1] G. Nicolis and I. Prigojin, *Self-Organization in Nonequilibrium Systems*, J. Wiley, New York (1977).
[2] L. P. Kubin, in *Material Science and Technology*, ed. by R. W. Cahn, P. Haase, and E. J. Kramer, VCH, D-Weinberg (1993), Vol. 6, p. 138.
[3] G. A. Malygin, Fiz. Tverd. Tela **37**, 3 (1995) [Solid State Phys. **37**, 1 (1995)].
[4] V. S. Bobrov, S. I. Zaĭtsev, and M. A. Lebyodkin, Fiz. Tverd. Tela **32**, 3060 (1990) [Sov. Phys. Solid State **32**, 1176 (1990)].
[5] M. A. Lebyodkin, Y. Bréchet, Y. Estrin, and L. P. Kubin, Phys. Rev. Lett. **74**, 4758 (1995).
[6] M. A. Lebyodkin, Y. Bréchet, Y. Estrin, and L. P. Kubin, Solid State Phenom. **42–43**, 313 (1995).
[7] M. A. Lebyodkin, Y. Estrin, Y. Bréchet, and L. P. Kubin, Acta Mater. **44**, 4531 (1996).
[8] A. Portevin and F. Le Châtelier, Trans. ASST**5**, 457 (1924).
[9] L. P. Kubin and Y. Estrin, Acta Metall. Mater. **38**, 697 (1990).
[10] A. A. Andronov, A. A. Vitt, and S. É. Khaikin, *Theory of Oscillations*, Fizmatgiz, Moscow (1959); transl. Pergamon, Oxford (1966), Chs. IV and IX.
[11] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. A **38**, 364 (1988).
[12] J. M. Carlson, J. S. Langer, and B. E. Shaw, Rev. Mod. Phys. **66**, 657 (1994).
[13] G. Ananthakrishna, Scr. Metall. Mater. **29**, 1183 (1993).
[14] G. Ananthakrishna, C. Fressengeas, M. Grosbras *et al.*, Scr. Metall. Mater. **32**, 1731 (1995).
[15] L. P. Kubin and Y. Estrin, in *Continuum Models for Materials with Microstructure*, ed. by H.-B. Muhlhaus, Wiley & Sons, New York (1995), p. 395.
[16] M. A. Lebyodkin and L. R. Dunin-Barkowskiĭ, Fiz. Tverd. Tela **40**, 487 (1998) [Solid State Phys. **40**, 447 (1998)].
[17] J. D. Eshelby, in *Progress in Solid Mechanics 2*, ed. by I. N. Sneddon and R. Hill, North-Holland, Amsterdam (1961), p. 89.
[18] P. Hähner, Mater. Sci. Eng., A **164**, 23 (1993).
[19] P. Penning, Acta Metall. **20**, 1169 (1972).
[20] H. M. Zbib and E. C. Aifantis, Res. Mech. **23**, 261 (1988).
[21] L. P. Kubin, K. Chihab, and Y. Estrin, Acta Metall. **36**, 2707 (1988).
[22] M. A. Lebyodkin, L. Dunin-Barkowskii, Y. Bréchet, L. P. Kubin, and Y. Estrin, Mater. Sci. Eng., A **234–236**, 115 (1997).
[23] J. H. E. Cartwright, E. Hernández-García, and O. Piro, Phys. Rev. Lett. **79**, 527 (1997).

Translation provided by the Russian Editorial office.