Stability of self-gravitating astrophysical disks

V. L. Polyachenko and E. V. Polyachenko

Institute of Astronomy, Russian Academy of Sciences, 109017 Moscow, Russia
(Submitted 26 November 1996)
Zh. Éksp. Teor. Fiz. **112**, 771–795 (September 1997)

Criteria which guarantee the stability of self-gravitating gaseous and stellar disks toward any localized small perturbations are obtained. These criteria are formulated as inequalities of the form $Q>Q_c$ (separately for gas and stars). The latter should be satisfied by the ''stability parameter'' Q, which is equal, by definition, to unity on the stability boundary of radial perturbations. The critical value of the stability parameter Q_c is appreciably greater than (although of the order of) unity, attesting to the great instability of nonaxially symmetric perturbations. It is shown that the stability criterion derived for gaseous disks is valid for disks rotating within a spheroidal component (as in spiral galaxies) or in the field of a central mass (planetary rings and accretion disks). Stellar disks are stabilized with significantly greater difficulty. This is attributable mainly to the anisotropy of the velocity distribution inherent to them, which is favorable for instability. © *1997 American Institute of Physics.* [S1063-7761(97)00109-1]

1. INTRODUCTION

The dynamic instabilities of gravitating systems are studied for several reasons. First of all, the development of instabilities can result directly in the formation of different observable structural features (such as spiral arms of galaxies, galactic bars, or a thin planetary ring structure). The ellipsoidal shape of elliptical galaxies also arises as a result of instability.[1] In some cases there is interest in equilibrium states that are similar to some other, simpler states with greater symmetry. As an example, we can cite the disk systems in the theory of spiral structures,[2,3] which examines nonaxially symmetric equilibria (of spiral shape) that differ slightly from the axially symmetric ''background.'' The problem of finding such equilibrium states can be regarded as a special case of the problem of small fluctuations. The stability requirement often imposes fairly rigid constraints on the parameters of equilibrium systems. This is clearly illustrated by the history of the interpretation of the anomalously rapid increase in the luminosity of the stars at the center of the spherical galaxy M87, which was discovered by Sargent *et al.*[4] (similar ''luminosity cusps'' were subsequently found in several other galaxies). The attempts to explain these data within standard models that are isotropic near the center were unsuccessful. Sargent *et al.*[4] suggested that there is a supermassive (with a mass of the order of $5\times10^9\ M_\odot$) black hole at the center of M87. Without assuming that the less exotic explanations had been exhausted, Duncan and Wheeler[5] considered the possibility of anisotropic equilibria and showed that the effect observed can also be explained, in principle, by a fairly great dominance of the radial velocities over the transverse velocities. This idea was subsequently developed in many papers by other investigators (see, for example, Ref. 6). However, it was found[3] that the necessary degree of anisotropy is so great that all the anisotropic models considered must be unstable. This assumption is attributable to the instability of radial orbits (for further details, see Ref. 1), which leads to the transformation of the originally spherical system into an elliptical system. Therefore, systems with the anisotropy adopted in Refs. 5 and 6 could not, in fact, remain spherical (in contradiction of the observed spherical shape of the isophotes).

Disk systems are highly dynamic and are subject to various instabilities. Spiral galaxies, accretion disks, and broad planetary rings (as around Saturn), as well as such hypothetical objects as protogalaxies and protoplanetary clouds, are classified as astrophysical disks. The strongly flattened, disk-shaped form of all these objects is due to their rapid rotation.[1)] Thus, equilibrium is established in a simple manner in such disks, i.e., it is governed mainly by the balance between the centrifugal and gravitational forces. However, the use of the simplest model, i.e., an infinitely thin disk with zero pressure, would be incorrect because of the strong instability of such a disk. This follows formally from the local dispersion equation derived by Toomre,[7] which relates the frequency ω and the wave number k_r of a short-wavelength axially symmetric (annular; see Fig. 1a) perturbation near a certain arbitrary radius r:

$$\omega^2=\kappa^2(r)-2\pi G\sigma_0(r)|k_r|, \tag{1}$$

where $\kappa(r)=\sqrt{4\Omega^2+r(\Omega^2)'}$ is the epicyclic frequency, $\Omega(r)$ is the angular velocity, $(\Omega^2)'\equiv d\Omega^2/dr$ (as a rule, the rotation is differential, i.e., $\Omega'\neq0$, and $\Omega'<0$), $\sigma_0(r)$ is the surface mass density of the disk, G is the gravitational constant, the perturbation is assumed to be proportional to $\exp(-i\omega t+ik_r r)$, and the disk is oriented in the xy plane. Such a form for the dispersion relation seems fairly straightforward. When $G=0$, i.e., when the self-gravitation of the disk is neglected, we have, according to (1), oscillations with a frequency κ, i.e., epicyclic oscillations, and the equation

$$\ddot{r}+\kappa^2 r_1=0$$

for a small perturbation r_1 of the radius of the initially circular orbit $[r(t)=r_0+r_1(t)]$ can easily be obtained by linearizing the equation of motion of a star

$$\ddot{r}=-\frac{dW}{dr},$$

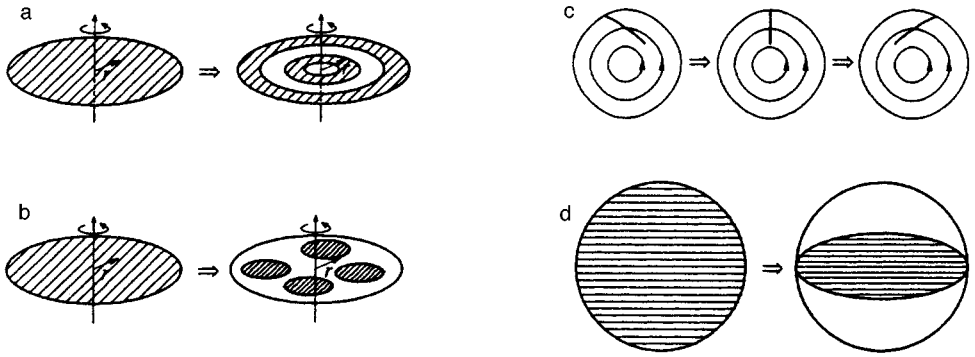

FIG. 1. Some examples of the evolution of self-gravitating disks: a—axially symmetric (radial) perturbations; b—evolution of an initially cold (with circular particle orbits) disk; c—winding of a spiral perturbation in a differentially rotating disk; d—bar mode (elliptical deformation) of a disk.

where the effective potential energy $W=\Phi_0+L_z^2/2r^2$ (Φ_0 is the gravitational potential, and $L_z=rv_\varphi=\text{const}$ is the angular momentum of the star). It is easy to prove that in a coordinate frame rotating at the rate $\Omega(r)$ the perturbed motion of particles takes place along small, azimuthally extended ellipses (epicycles) with the aspect ratio

$$b/a=2\Omega(r)/\kappa(r)>1. \tag{2}$$

As is seen from Eq. (1), self-gravitation leads to the instability of fairly small-scale perturbations: $\omega^2<0$, if $|k_r|>k_T\equiv\kappa^2/2\pi G\sigma_0$. Of course, this instability is physically Jeans instability (for further details on Jeans instability and its applications, see Refs. 2, 3, and 8). The square of the ''ordinary'' Jeans frequency $\omega_{J3}^2=4\pi G\rho_0(r)$ [which should have appeared on the right-hand side of Eq. (1) for a medium that is infinite along z, i.e., a cylinder with a bulk density $\rho_0(r)$] is replaced in the disk case by $\omega_{J2}^2=2\pi G\sigma_0(r)|k_r|$. The need for just such a replacement follows from dimensionality arguments: The square of the frequency can be constructed in only one way from the three parameters at our disposal G, σ_0, and k_r; i.e., $\omega_{J2}^2=\text{const}\cdot G\sigma_0(r)|k_r|$, and the value $\text{const}=2\pi$ corresponds to the fact that ω_{J2}^2 transforms into ω_{J3}^2 when $k_r=2\pi/h$ (h is the half-thickness of the disk).

Equation (1) can naturally be generalized to the dispersion relation (see, for example, Refs. 2 and 3)

$$\omega^2=\kappa^2(r)-2\pi G\sigma_0(r)|k_r|+k_r^2c_s^2, \tag{3}$$

for radial perturbations of a thin disk with a finite two-dimensional pressure $P_\perp=\int P\,dz$ and $P_z=0$, where P is the ordinary gas pressure, c_s is the velocity of sound, and $c_s^2=\partial P_\perp/\partial\sigma_0$. As expected, when k_r is large, Eq. (3) degenerates to the dispersion relation for sound waves. Equation (3) can be represented in dimensionless form as

$$\overline{\omega^2}=1-|\overline{k}|+\overline{k^2}Q^2/4, \tag{4}$$

where $\overline{\omega}=\omega/\kappa$, $\overline{k}=k_r/k_T$, and

$$Q=\kappa c_s/\pi G\sigma_0. \tag{5}$$

The behavior of the function $\overline{\omega^2}=\overline{\omega^2}(\overline{k})$ for various values of Toomre's stability parameter[2)] Q is shown in Fig. 2. The parameter Q is a combination of equilibrium parameters of the disk: the parameters κ and c_s, which promote stabilization of the disk as they increase, appear in the numerator, while σ_0, which promotes destabilization as it increases, appears in the denominator. Marginal curve *2*, which is tangent to the horizontal axis, corresponds to $Q=1$; when $Q>1$, we have curves of type *1*, on which $\omega^2>0$ (stability) exists everywhere, and when $Q<1$, we have curves of type *3*, for which there is an interval (k_1,k_2) of unstable wave numbers.

The principal factors regarding the stability of gravitating disks toward radial perturbations are as follows: in this case stability or instability is determined by the value of the single dimensionless parameter Q. Let us now turn to perturbations which destroy the original axial symmetry of a disk, i.e., which are proportional to $\exp(-i\omega t+ik_r r+im\varphi)$, where φ is the azimuth, and m is an integer ($m\neq 0$). These perturbations, in contrast with annular perturbations, are subject to the influence of the differential character of the rotational motion. Figure 1c schematically illustrates the evolution of a spiral perturbation as a result of differential rotational motion. This evolution pattern is associated with significant complications in the stability theory for nonaxially symmetric perturbations (see below). On the other hand, nonaxial perturbations are more unstable (and, therefore, more interesting) than radial perturbations; more precisely, they are less amenable to stabilization. This, in particular, has been graphically demonstrated by numerous computer (N-body) experiments. For example, the evolution of an initially ''cold'' disk (with circular orbits for all the particles) has been simulated repeatedly (see, for example, Ref. 9); the result of this evolution is schematically represented in Fig. 1b: the disk breaks up into several fragments (which subsequently coalesce to form an elliptical disk with an appreciable percentage of radial particle motion). As we can see, other than annular perturbations dominate here. The dominant role of the nonradial perturbations is manifested even more graphically in numerical experiments with disks on

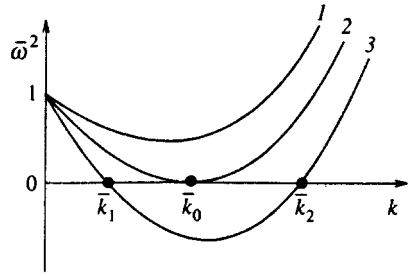

FIG. 2. Dispersion curves for various disk perturbations.

which the condition for stability toward radial perturbations, i.e., $Q(r)\equiv 1$, holds at each point. It has been found[10] that such disks remain unstable with respect to nonradial modes (Fig. 1d).

The complications arising in the construction of a stability theory for nonaxisymmetric perturbations in a differentially rotating, gravitating disk are akin in many respects to problems known, for example, in the hydrodynamic theory of stability. We shall focus on the local solutions near a certain radius r_0, representing the angular velocity as $\Omega(r)=\Omega(r_0)+\Omega'(r_0)(r-r_0)$ and omitting the remaining terms of the expansion in a Taylor series. Such an approximation is similar to the Couette approximation in the case of plane-parallel flow, $v=v_x=v_x(y)$, where the linear law $v_x(y)=v_x(y_0)+v_x'(y_0)(y-y_0)$ is adopted. In the case of incompressible Couette flow, the linearized problem is known to reduce to Rayleigh's equation,[11] and it then becomes impossible to satisfy the necessary conditions on the flow boundaries (or in the limit $|y|\to\infty$).[12] Therefore, to study the dynamics of the perturbations in such flows, an initial-value problem must be solved. A similar situation arises when local perturbations are considered in an incompressible rotating flow.[13] It is easy to show that in the x representation this problem also reduces to the Rayleigh equation. Hence it follows that there are no characteristic solutions which vanish far from the location of a perturbation because of the impossibility of satisfying the boundary conditions.

In the differentially rotating disks that we considered local nonaxisymmetric irrotational[3)] perturbations likewise cannot be characteristic in an approximation similar to the Couette approximation, i.e., under a linear law for the angular velocity $\Omega(r)$ (see Sec. 2.2 below).[4)] To analyze the dynamics of a disk, an initial-value problem must be solved in the general case. However, the available information on the state and structure of the disk and on the initial level and character of the perturbations is generally insufficient for such an analysis. In this situation approximate criteria, which define the possibility of an appreciable increase in the initial perturbations and which depend only on the basic parameters of the disk, can be of great value. Below we propose a procedure for obtaining such criteria (which are similar to Toomre's criterion) for local nonaxisymmetric disk perturbations. More precisely, we shall limit the discussion to the derivation of criteria which establish the impossibility of the quasiexponential growth of perturbations. We note, however, that a strong increase in the initial perturbation amplitude occurs only in the presence of a sufficiently prolonged period of quasiexponential growth (for further details, see the discussion of Fig. 4 in Secs. 2 and 3 and especially in the concluding remarks).

It turns out that just as the inequality $Q>1$ ensures the stability of a disk toward axially symmetric perturbations, arbitrary local perturbations are guaranteed not to grow, if the following similar, but stronger inequality holds:

$$Q>Q_c>1, \tag{6}$$

where Q_c is a new critical value of Toomre's parameter.

The main purpose of the present work is specifically to calculate Q_c for gaseous (Sec. 2) and stellar (Sec. 3) disks. For gaseous disks we use two models that have long been considered standard: polytropic disks of finite thickness (Sec. 2.1) and an infinitely thin disk with a two-dimensional hydrodynamic pressure (Sec. 2.2). An expression for the critical value $Q_c^{(g)}$, which is a function of the parameter $\alpha^2=2\Omega(r)/r|\Omega'(r)|$ that characterizes the degree of differential rotation of the disk at the point r [see Eq. (18) below], is found in Sec. 2.2. This expression remains the same for disks with an arbitrary value of the adiabatic exponent γ. The results of this section were contained to a considerable degree in a veiled form in Goldreich and Lynden-Bell's work,[14] where the stability of gaseous disks with differential rotation was investigated for the first time. Real astrophysical gaseous disks (see the beginning of the paper) either occupy the equatorial plane within a spheroidal component in a spiral galaxy or rotate in the field of a central mass (as do planetary rings or accretion disks); the equilibrium state is determined not only by self-gravitation, but also by the gravitational force outside of the disk. However, in Goldreich and Lynden-Bell's theory this external force is taken into account not entirely correctly (for further details see Sec. 2.1). Therefore, Goldreich and Lynden-Bell's theory must be modified to take into account the real situation. A very simple modification is considered in Sec. 2.2, where we investigate the stability of a model of an infinitely thin disk with a two-dimensional pressure $P_\perp$. Here the part of the equilibrium which is provided by the self-gravitation and the part which is provided by the masses outside of the disk become indistinguishable. It is noteworthy that the stability criterion (6) with the same $Q_c^{(g)}(\alpha^2)$ as for polytropic disks remains in force in this case. Because this criterion remains unchanged when very different models are used, it can be presumed that it is universal: it is apparently valid for any gaseous disk. In Sec. 3 we analyze the local stability of stellar disks (in the approximation which Julian and Toomre[15] previously used to investigate the evolution of initial perturbations). It was shown that the integral equation for the Fourier amplitude of the potential has solutions, whose behavior is qualitatively similar to the behavior of the perturbations in a gaseous disk. An approximate equation specifying the stability boundary of a stellar disk is obtained. A numerical investigation of this equation makes it possible to find the function $Q_c^{(s)}(\alpha^2)$, and it turns out that the critical values $Q_c^{(s)}(\alpha^2)$ are significantly greater than the corresponding $Q_c^{(g)}(\alpha^2)$. The reasons why it is considerably more difficult to stabilize stellar disks than gaseous disks are considered. A discussion of the results can be found in Sec. 4.

2. GASEOUS DISKS

2.1. Stability of polytropic disks of finite thickness

The first investigation of the behavior of local perturbations in differentially rotating, self-gravitating disks of finite thickness was performed by Goldreich and Lynden-Bell.[14] The approach adopted by those investigators to the problem is as follows.

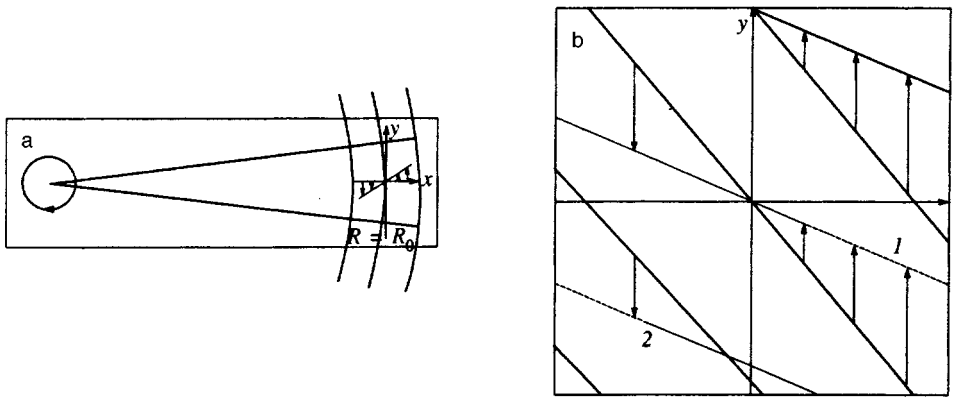

FIG. 3. Rotation with uniform displacement of the angular velocity (a) and accompanying Goldreich-Lynden-Bell coordinates (b): *1*—linear plot of $y'(\tau+\delta\tau)=\text{const}$; *2*—linear plot of $y'(\tau)=\text{const}$; the arrows correspond to the value of $2Axdt$.

First of all, they introduce (Fig. 3) the coordinate axes x', y', and z', which are associated with the unperturbed gas flow:

$$x'=x, \quad y'=y-2Axt, \quad z'=z, \tag{7}$$

where x, y, and z define the local Cartesian coordinate frame at a given point on the disk; the x axis is directed along the radius, the y axis is directed along the azimuth, and the z axis is parallel to the ratio axis of the disk; and $A=r\Omega'(r)/2$. As a result, the problem reduces to the solution of a system of linearized hydrodynamic equations and Poisson's equation with coefficients that depend on t and z, but do not depend on x' and y'. At the same time, in the original untransformed equations the coefficients depend on x and z, but not on t and y. Thus, the transformation (7) permits the transition from inhomogeneity with respect to x to inhomogeneity with respect to t. This makes it possible to correctly examine the problem of the temporal evolution of perturbations that are proportional to $\exp(ik_x x'+ik_y y')$ with the arbitrary constants k_x and k_y. The inverse transformation of this exponential function to the original (x,y,z) frame gives

$$\exp(ik_x x'+ik_y y')=\exp[ik_y(-\tau x+y)], \tag{8}$$

where we have introduced the new ''time'' variable

$$\tau=2At-k_x/k_y;$$

$\tau=0$ corresponds to the case in which the wave vector of the perturbation in the (x,y,z) frame is exactly parallel to the y axis (the middle drawing in Fig. 1c). The right side of the figure shows that the radial component of the wave vector (in the original coordinate system) varies with time according to a linear law:

$$k_r=-k_y\tau=k_x-2Ak_y t. \tag{9}$$

Next, making additional assumptions,[5)] Goldreich and Lynden-Bell derived the evolution equations for $\theta_1^*\equiv\sigma_1/\sigma_0$ (σ_1 and σ_0 are, respectively, the perturbed and unperturbed surface densities) in two cases, viz., for disks with the adiabatic exponents $\gamma=2$ and $\gamma=1$. In the case of an isometric disk ($\gamma=1$) the equation for nonradial perturbations in the notation of the original work in Ref. 14 is

$$\frac{d}{dt}\left(\frac{\theta_1^*}{1+\tau^2}\right)+\left[\frac{2B/A}{(1+\tau^2)^2}+\frac{B\Omega/A^2-P'/g(\overline{m})}{1+\tau^2}\right]\theta_1^*=0, \tag{10}$$

where

$$P'=\pi G\rho_c/A^2,$$

$$\frac{1}{g(\overline{m})}=\frac{\overline{m}(1-\overline{m^2})}{1+\overline{m}+(1/2)\overline{m}^2\Psi'(1+\overline{m/2})}, \tag{11}$$

$$\Psi'\left(1+\frac{\overline{m}}{2}\right)=\sum_{r=1}^{\infty}\frac{1}{(\overline{m/2}+r)^2},$$

$$\overline{m}=k_y\sqrt{1+\tau^2}/k_0, \quad B=A+\Omega, \quad k_0^2=2\pi G\rho_c/c^2,$$

and ρ_c is the density of the disk in the $z=0$ equatorial plane. In this model the bulk density $\rho_0(z)=\rho_c/\cosh^2(z/z_0)$, where $z_0=c_s/\sqrt{2\pi G\rho_c}$. The analogous equation for the evolution of radial perturbations is somewhat simpler:

$$\frac{d^2\theta_1^*}{d\tau^2}+\left[\frac{B\Omega}{A^2}-\frac{P'}{g(m_x)}\right]\theta_1^*=0, \tag{12}$$

where $m_x=k_x/k_0$. In the case of a disk with $\gamma=2$, the evolution equations nearly coincide with (10) and (12), except that, instead of $P'/g(\overline{m})$, they contain a somewhat different function $P/F(K)$. Here P is a constant that is analogous to P', $F(K)$ is a function that replaces $g(\overline{m})$, and $K=k_y a\sqrt{1+\tau^2}$, where a is the half-thickness of the disk. The explicit form of this function, as well as the constants P, are not needed further. It is important that both functions, $1/g(m_x)$ and $1/F(K)$, have a maximum at a certain value of m_x or K, respectively. For this reason, it follows from Eq. (12) (and the analogous equation for $\gamma=2$) that under the condition

$$\frac{B\Omega}{A^2}\geqslant\begin{cases}P'\ \max[1/g(m_x)], & \gamma=1,\\ P\ \max[1/F(K)], & \gamma=2\end{cases} \tag{13}$$

radial perturbations are stable. Selecting the equality sign in the lower relation in (13) (disks with $\gamma=2$), Goldreich and Lynden-Bell numerically solved an equation like (10) for nonaxially symmetric perturbations in disks near the stability boundary of the radial modes. The solutions which we obtained appear to be approximately as shown in Figs. 4a and

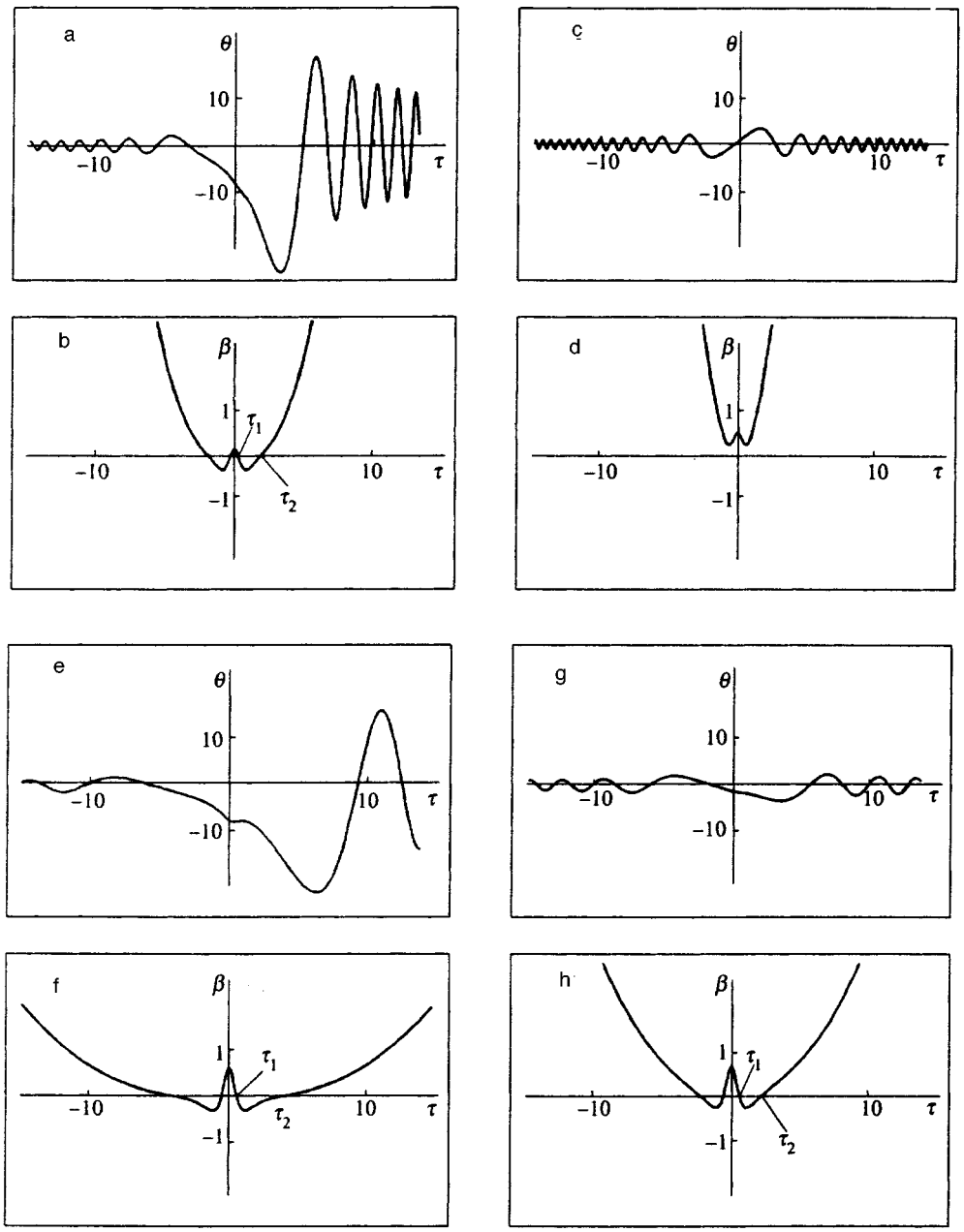

FIG. 4. Evolution of nonaxially symmetric perturbations in gaseous disks: a—$Q^2=1.5$, $\alpha^2=2$ (galaxy with a flat rotation curve); b—the function $\beta(\tau)$ corresponding to case a; c—$Q^2=4$, $\alpha^2=2$; d—the function $\beta(\tau)$ corresponding to case c; e—$Q^2=1.5$, $\alpha^2=4/3$ (an accretion disk with Keplerian rotation); f—the function $\beta(\tau)$ corresponding to case e; g—$Q^2=4$, $\alpha^2=4/3$; h—the function $\beta(\tau)$ corresponding to case g. The perturbations in a stellar disk behave in a similar manner; however, these perturbations are stabilized at significantly greater values of Q.

4e: at first we have oscillations with a small amplitude $(\theta_1^*)_i$, and then, after a certain time τ_1, the perturbation grows (especially rapidly near $\tau=0$), and finally, after $\tau=\tau_2$, oscillations with a significantly increased amplitude $(\theta_1^*)_f$ are established; the enhancement factor $(\theta_1^*)_f/(\theta_1^*)_i$ can reach values of the order of 10^3.

Goldreich and Lynden-Bell did not address the question of the existence of a certain condition which guarantees complete stability of a disk. Let us try to extract such information from Eqs. (10) and (12). We first turn to Eq. (12), which describes the radial perturbations of an isothermal disk. Introducing the notation $\max[1/g(m_x)]=c_1$, we write the condition of marginal stability in the form

$$\frac{\kappa^2}{4\pi G\rho_c}=c_1. \tag{14}$$

As follows from the explanations and the data in the Introduction, it is natural to formulate Toomre's stability parameter Q as an equality to unity. In this case we clearly have

$$Q=\frac{\kappa}{\sqrt{4\pi G\rho_c c_1}}, \tag{15}$$

since, first, the parameter Q thus defined is proportional, as it should be, to $\kappa c_s/\pi G\sigma_0$ (since $\sigma_0\sim\rho_c h$, and $c_s/h\sim\sqrt{4\pi G\rho_c}$ because the equilibrium along z is assumed to be the same as in an isolated disk) and, secondly, since $Q=1$ on the stability boundary of axially symmetric perturbations according to (14). We note that in the case at hand Q depends on only two independent parameters (κ and ρ_c). In the general case (5) there are three such parameters (κ, c_s,

and σ_0). The additional degree of freedom is clearly provided by the arbitrary nature of the mass ratio between the halo and the disk.

Let us now move on to nonaxial perturbations. Using the replacement $\theta_1^* = \sqrt{1+\tau^2}\,\theta(\tau)$ to eliminate the first derivative from Eq. (10), we reduce it to the equation

$$\ddot{\theta} + \beta(\tau)\,\theta = 0, \tag{16}$$

where the function $\beta(\tau)$ can be brought into the form

$$\beta(\tau) = (\alpha^4 - \alpha^2) - \frac{2\alpha^2}{1+\tau^2} + \frac{3}{(1+\tau^2)^2} - \frac{(\alpha^4-\alpha^2)}{Q^2 c_1 g(\overline{m})}. \tag{17}$$

Here $\alpha^2 = 2\Omega(r)/r|\Omega'(r)|$. The equality $\beta(\tau)=0$ separates the quasiexponentially growing (for $\beta<0$) and oscillating (for $\beta>0$) portions of the solutions $\theta(\tau)$ of Eq. (16). Let α^2 be fixed. For example, let $\alpha^2=2$ in the case of a flat rotation curve, where $V_\varphi = \Omega(r)r = \text{const}$. Such rotation is characteristic of most spiral galaxies. As can be shown, when $Q^2 \approx 1$ (as in Goldreich and Lynden-Bell's calculations), $\beta(\tau)$ vanishes at two points (see Fig. 4). However, at larger values of Q^2 (and $\alpha^2 > 3/2$) $\beta(\tau)$ is positive everywhere, as is seen from (17), so that perturbations never grow. This means that the function $Q^2_{\alpha^2} = Q^2_{\alpha^2}(\tau^2, \overline{m})$, which is defined according to (17) on the basis of the condition $\beta=0$, has a maximum (which depends only on α^2) as a function of two variables (τ^2 and $\overline{m}$). This maximum also specifies the critical value sought $Q = Q_c$, which guarantees complete stability of a disk (when $Q \geqslant Q_c$):

$$Q_c^2 = \frac{3}{2}\,\frac{\alpha^2-1}{\alpha^2-3/2}. \tag{18}$$

Several remarks should be made regarding Eq. (18).

1. The calculation of Q_c^2 does not require knowledge of the specific value (c_1) of the maximum of the function $P'/g(\overline{m})$: It is sufficient to know only that this maximum exists. In fact, the stability toward radial perturbations is determined, according to (12), by the same function $P'/g(m_x)$.

2. For the same reason, Q_c also has the form (18) for disks with $\gamma=2$.

3. In addition, it can be shown that the same expression (18) is valid for a disk with any adiabatic exponent γ. Although evolution equations can be obtained analytically[14] only for $\gamma=1$ and $\gamma=2$, the function $P_\gamma/F_\gamma(K)$, which replaces $P/F(K)$ when $\gamma=2$ or $P'/g(\overline{m})$ when $\gamma=1$, can be calculated numerically for any γ. This was done in Ref. 17, where it was also shown that the functions $P_\gamma/F_\gamma(K)$ exhibit qualitatively identical behavior for all γ; in particular, they all have a maximum at a certain value of the argument. Moreover, the existence of a maximum on any $P_\gamma/F_\gamma(K)$ curve is clearly seen from simple physical arguments. In fact, the square of the frequency of the radial perturbations at the point r is

$$\omega^2 = \kappa^2 - r^2(\Omega')^2\,\frac{P_\gamma}{F_\gamma(k_x)}. \tag{19}$$

As was noted above, radial perturbations, in contrast with nonradial perturbations, have a purely Jeans nature. However, this means that ω^2 passes, as, for example, on the curves in Fig. 2, through a minimum [and the function $P_\gamma/F_\gamma(k_x)$ accordingly passes through a maximum] at $k_x h \sim 1$, where h is of the order of the disk thickness. In the case of nonaxially symmetric perturbations the mechanism of gravitational (Jeans) compression is no longer the only cause of instability: another mechanism, which is associated with redistribution of the angular momentum of the system, acts here. This mechanism was examined in detail, for example, by Lynden-Bell and Kalnajs[18] (see also Refs. 2 and 3); the appearance of the terms that depend on τ in Eq. (17) is associated with it. Thus, it has been shown in this section that a stability criterion like (6) with the critical value Q_c given by the universal formula (18) is valid for disks with any adiabatic exponent.

4. We cannot ignore the presence of a singularity in Eq. (18): $Q_c^2 \to \infty$ when $\alpha^2 \to 3/2$. This circumstance means that there is a certain region of quasiexponential growth of nonaxisymmetric perturbations in disks having a decrease in the angular velocity with the radius at any Q that is faster than $\Omega \propto r^{-4/3}$. This finding can also be regarded, in a certain sense, as evidence of the great instability of disks with a strong degree of differential rotation. In fact, however, at large Q the width of the region of quasiexponential growth is very narrow, and its presence has virtually no influence on the evolution of the perturbation. We shall address this question again in the Conclusions.

2.2. Stability of an infinitely thin disk with a two-dimensional hydrodynamic pressure

The starting system consists of the equations of two-dimensional hydrodynamics and Poisson's equation:

$$\begin{aligned} &\frac{\partial \sigma}{\partial t} + \operatorname{div}(\sigma \mathbf{v}) = 0, \\ &\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\nabla)\mathbf{v} = -\nabla\Phi - \frac{1}{\sigma}\nabla P_\perp, \\ &\Delta\Phi = 4\pi G\sigma\delta(z), \end{aligned} \tag{20}$$

where σ is the surface mass density of the gas in the disk, $\mathbf{v}$ is the velocity of a gas element in the (x,y) plane, Φ is the gravitational potential, and $P_\perp$ is the pressure of the two-dimensional gas. Writing the linearized hydrodynamic equations (20) in the polar coordinates r and φ, we obtain

$$\begin{aligned} &\omega_* \mu - \frac{m}{r} v_\varphi + \frac{i}{\sigma_0}\frac{\partial}{\partial r}(\sigma_0 v_r) + \frac{i}{r} v_r = 0, \\ &\omega_* v_r - 2i\Omega v_\varphi + i\frac{\partial S}{\partial r} = 0, \\ &\omega_* v_\varphi + i\frac{\kappa^2}{2\Omega} v_r - \frac{m}{r} S = 0. \end{aligned} \tag{21}$$

Here it is assumed that the perturbed parameters are proportional to $\exp(-i\omega t + im\varphi)$ (m is a positive integer); $\mu = \sigma_1/\sigma_0$, σ_1 and σ_0 are, respectively, the perturbed and unperturbed surface densities; v_r and v_φ are the components of the perturbed velocity; $S = \Phi_1 + c_s^2\mu$; Φ_1 is the perturbed

gravitational potential; and $\omega_* = \omega - m\Omega = m(\Omega_p - \Omega) + i\gamma$, where γ denotes the small growth rate of the wave, and Ω_p is its angular velocity.

We introduce a local Cartesian coordinate frame near the corotation radius $r = r_0$ $[\Omega(r_0) - \Omega_p]$: $x = (r - r_0)/\Delta$, $\Delta = \kappa/m|\Omega'|$. In the first order with respect to the small parameter $\Delta/r_0 \ll 1$ (which is equivalent to $m \gg 1$[6]) Eqs. (21) take the form

$$
\begin{aligned}
&z\mu - \frac{k_\varphi}{\kappa} v_\varphi + \frac{i}{\kappa\Delta} + \frac{i}{r} v_r' = 0,\\
&z v_r - i\frac{2\Omega}{\kappa} v_\varphi + \frac{i}{\kappa\Delta} S' = 0,\\
&z v_\varphi + i\frac{\kappa}{2\Omega} v_r - \frac{k_\varphi}{\kappa} S = 0,
\end{aligned}
\tag{22}
$$

where $z = \omega_*/\kappa = \mathrm{Re}\, z + i\nu$, and $k_\varphi = m/r_0$. Using the first equation in (22) to eliminate v_φ from the remaining equations of this system, we obtain

$$
\begin{aligned}
&z v_r + \frac{\alpha^2}{k^2} v_r' - i\frac{\alpha^2}{k^2} z\kappa\Delta\mu + \frac{i}{\kappa\Delta} S' = 0,\\
&i\frac{k^2}{\alpha^2} v_r + i z v_r' + z^2\kappa\Delta\mu - \frac{k^2}{\kappa\Delta} S = 0,
\end{aligned}
\tag{23}
$$

where $\alpha^2 = 2\Omega_0/r_0|\Omega_0'|$, and, as it is easy to show, $k^2 = \alpha^4 - \alpha^2$.

In (23) we move over to the Fourier transforms of the perturbed functions:

$$
A_q = \int_{-\infty}^{\infty} dx e^{-iqx} A(x). \tag{24}
$$

In q space Eqs. (23) take the form

$$
\begin{aligned}
&i\left(\frac{d}{dq} + \nu\right) v_{rq} + i\frac{\alpha^2}{k^2} q v_{rq}\\
&\quad + \frac{\alpha^2}{k^2}\kappa\Delta\left(\frac{d}{dq} + \nu\right)\mu_q - \frac{q}{\kappa\Delta} S_q = 0,\\
&-i\left(\frac{d}{dq} + \nu\right) q v_{rq} + i\frac{k^2}{\alpha^2} v_{rq}\\
&\quad - \kappa\Delta\left(\frac{d}{dq} + \nu\right)^2 \mu_q - \frac{k^2}{\kappa\Delta} S_q = 0.
\end{aligned}
\tag{25}
$$

We make the change in notation $A_q \exp(q\nu) \to A_q$, which is equivalent to writing Eqs. (25) without ν. From (25) we find

$$
v_{rq} = \frac{i\kappa\Delta}{2 - \alpha^2(q^2 + k^2)/k^2}\left[\mu_q'' - \frac{\alpha^2}{k^2} q\mu_q' + \frac{q^2 + k^2}{\kappa^2\Delta^2} S_q\right]. \tag{26}
$$

Differentiating the latter expression and substituting the result into Eq. (25), we obtain the following equation, which relates μ_q and S_q:

$$
\begin{aligned}
&\mu_q''' + \frac{\alpha^2}{k^2} q\frac{4 - \alpha^2(q^2 + k^2)/k^2}{2 - \alpha^2(q^2 + k^2)/k^2}\left(\mu_q'' - \frac{\alpha^2}{k^2} q\mu_q' + \frac{q^2 + k^2}{\kappa^2\Delta^2}\right) S_q\\
&\quad - \frac{\alpha^2}{k^2} q\mu_q'' - \frac{\alpha^2}{k^2}\mu_q' + \frac{q^2 + k^2}{\kappa^2\Delta^2} S_q' + \frac{2q}{\kappa^2\Delta^2} S_q\\
&\quad + \left[2 - \frac{\alpha^2}{k^2}(q^2 + k^2)\right]\left(\frac{q}{\kappa^2\Delta^2} S_q - \frac{\alpha^2}{k^2}\mu_q'\right) = 0.
\end{aligned}
\tag{27}
$$

It follows from Poisson's equation in the approximation of the small parameter Δ/r_0 that

$$
\mu_q = -\frac{\Phi_q\sqrt{q^2 + k^2}}{2\pi G\sigma_0\Delta}, \tag{28}
$$

so that

$$
S_q = \Phi_q(1 - a\sqrt{q^2 + k^2}),
$$

where $a = cc_s^2/2\kappa^2\Delta^2$, and $c = \kappa^2\Delta/\pi G\sigma_0$. We can then write Eq. (27) in the form

$$
a_0\Phi_q + a_1\Phi_q' + a_2\Phi_q'' + a_3\Phi_q''' = 0
$$

or, using the notation $a_i = a_3 b_i$ and $a_3 = -c\sqrt{q^2 + k^2}/2$, in the form

$$
b_0\Phi_q + b_1\Phi_q' + b_2\Phi_q'' + b_3\Phi_q''' = 0, \tag{29}
$$

where

$$
\begin{aligned}
b_0 = q\Bigg\{ &\frac{2}{q^2 + k^2 - 2k^2/\alpha^2}\left[\frac{\alpha^2}{k^2} - \frac{\alpha^2}{q^2 + k^2} - \frac{k^2}{(q^2 + k^2)^2}\right.\\
&\left. + \frac{2}{c}\sqrt{q^2 + k^2}(1 - a\sqrt{q^2 + k^2})\right] - \frac{3k^2}{(q^2 + k^2)^3}\\
&- \frac{\alpha^2}{k^2}\frac{3 - \alpha^2}{q^2 + k^2} - \frac{8}{c}\frac{1}{\sqrt{q^2 + k^2}}(1 - a\sqrt{q^2 + k^2}) + 2\frac{a}{c}\Bigg\},
\end{aligned}
\tag{30}
$$

$$
\begin{aligned}
b_1 = &1 - \frac{2}{c}\sqrt{q^2 + k^2}(1 - a\sqrt{q^2 + k^2})\\
&+ \frac{3k^2}{(q^2 + k^2)^2} - \frac{2\alpha^2}{q^2 + k^2},
\end{aligned}
\tag{31}
$$

$$
b_2 = q\left(\frac{3}{q^2 + k^2} - \frac{2}{q^2 + k^2 - 2k^2/\alpha^2}\right). \tag{32}
$$

It can be verified by direct substitution that Eq. (29) with the coefficients (32) can be factored in the form

$$
\left(\frac{d}{dq} + A\right)\left(\frac{d^2}{dq^2} + B\right)\Phi_q = 0, \tag{33}
$$

where $A = b_2$ and $B = b_1$. Such factoring, of course, is not accidental: it corresponds to separation of the vortex and gravitational-sound branches of the disk modes. We are interested only in the latter, since just this factor is a generalization of the Jeans branch for radial perturbations, which was studied by Toomre (see the Introduction); it corresponds to the solutions of Eq. (33) which obey the equation

$$
\left(\frac{d^2}{dq^2} + B\right)\Phi_q = 0, \tag{34}
$$

where

$$B(q)=1-\frac{2\alpha^2}{q^2+k^2}+\frac{3k^2}{(q^2+k^2)^2}-\frac{2}{c}\sqrt{q^2+k^2}$$
$$\times(1-a\sqrt{q^2+k^2}). \tag{35}$$

Equation (34) can be regarded as a Schrödinger equation with the potential energy $U(q)=-B(q)$ and the energy $E=0$. It is easy to see that the behavior of the function $U(q)$ is such that there are no ''good'' solutions of this equation which satisfy the natural conditions of decay as $|q|\to\infty$ (because there is no well); the only exception is the case of a cold disk ($c_s=0$). Therefore, we shall turn to an examination of the problem (in the spirit of Goldreich and Lynden-Bell's approach) of the evolution of a perturbation of form (8). This transition is accomplished by performing the substitution

$$k_r\Delta=q=-k_\varphi\Delta\tau, \tag{36}$$

where $k_\varphi=m/r$, and $\tau=r\Omega_0't-k_{r_0}/k_{\varphi,0}$ [this substitution has already been encountered above; see Eq. (9)]. Equation (34) takes the form

$$\frac{d^2\Phi}{d\tau^2}+\left[k_\varphi^2\Delta^2-\frac{2\alpha^2}{1+\tau^2}+\frac{3}{(1+\tau^2)^2}-\frac{2\sqrt{1+\tau^2}(k_\varphi\Delta)^3}{c}\right.$$
$$\left.\times(1-a\sqrt{1+\tau^2}k_\varphi\Delta)\right]\Phi=0. \tag{37}$$

We note that this equation could also be derived by the method used by Goldreich and Lynden-Bell, but an alternative derivation is proposed above. The first three terms in the square brackets coincide with the corresponding terms in Eq. (17), which was obtained by Goldreich and Lynden-Bell[14] for Cauchy's problem of the evolution of a perturbation of the form $\exp(-ik_\varphi\tau x+ik_\varphi y)$ (in a local Cartesian coordinate frame). This coincidence is, of course, not accidental. The fact that the problem can be reduced in the approximation under consideration to the equation for one q harmonic (34) in itself calls for such a coincidence. The last two terms in Eq. (37) differ from the terms corresponding to them in Goldreich and Lynden-Bell's equation (17), since they are specific to the model of an infinitely thin disk. Equation (37) can be rewritten as

$$\frac{d^2\Phi}{dt^2}+4A^2B(q(t))\Phi=0,\quad A=\frac{r_0}{2}|\Omega_0'|. \tag{38}$$

We see that the behavior of the solution corresponds to the sign of B: it oscillates (is stable) when $B>0$ and has a growing mode (is unstable) when $B<0$. The substitution

$$\frac{d^2}{dt^2}\to-\omega_*^2=-(\omega-m\Omega_0)^2$$

in Eq. (38) leads to a local dispersion relation, and the equation $B(q)=0$ is the condition for the absence of quasiexponentially growing perturbations. This condition can be written in the form

$$Q^2=4\left(\overline{\lambda}-\overline{\lambda^2}+g^2\overline{\lambda^4}-\frac{3k^2}{4\alpha^4}g^4\overline{\lambda^6}\right), \tag{39}$$

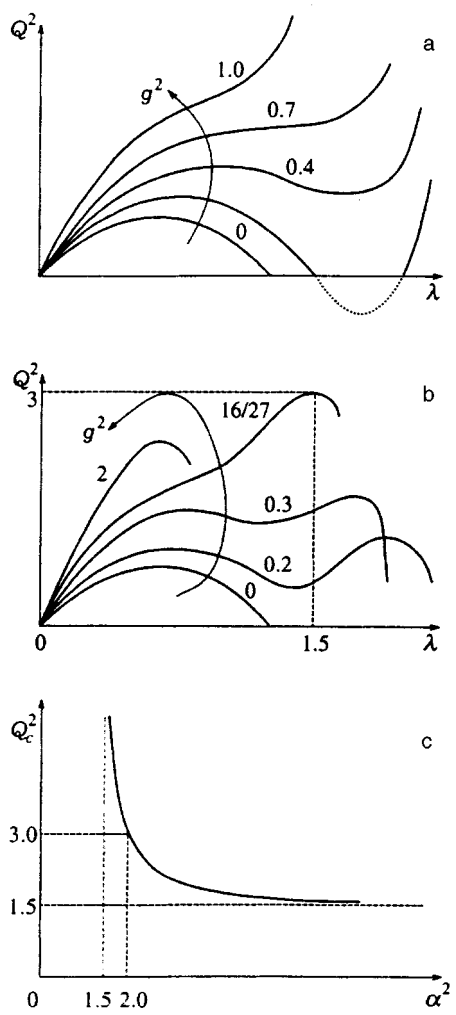

FIG. 5. Plots of the marginal stability $Q^2(\bar{\lambda})$ for the hydrodynamic model of a disk with a flat rotation curve ($\alpha^2=2$) and a two-dimensional pressure according to the work of Lin and Lau[19] or Bertin[20] (a) and according to our work (b), as well as dependence of the critical value $Q_c^2(\alpha^2)$ for the maximum of the function $Q^2_{\alpha^2}(\bar{\lambda},g^2)$ as a function of the two variables (c).

where $\bar{\lambda}=\lambda/\lambda_c$, $\lambda=2\pi/\sqrt{k_\varphi^2+k_r^2}$, $\lambda_c=4\pi^2G\sigma_0/\kappa^2$ is Toomre's critical wavelength, and $g^2=8\alpha^2/c^2=8r_0|(\Omega_0^2)'|k_\varphi^2(\pi G\sigma_0)^2/\kappa_0^6$ (we note here that $g^2\sim m^2$). Equation (39) defines the function $Q^2(\lambda,g)$ for an assigned α^2 [i.e., for a fixed rotation curve $\Omega=\Omega(r)$]. Plots of $Q^2(\bar{\lambda})$ for $\alpha^2=2$ and several values of g (which can be used to take into account the nonaxial character of a perturbation, if $k_\varphi=m/r\neq0$) are presented in Fig. 5b; when $g=0$, we have the marginal curve for axially symmetric disk perturbations. All these curves have maxima at $\bar{\lambda}=\bar{\lambda}_m$ for all α^2 and g^2. Moreover, the maxima for a given α^2 at first increase with increasing g^2 and then decrease (Fig. 5b). Therefore, the function $Q^2_{\alpha^2}(\lambda,g^2)$ has a (single) maximum as a function of the two variables, λ and g ($\lambda>0$, $g>0$). It follows from the equality $\partial Q^2/\partial g^2=0$ that $g^2\lambda^2=2\alpha^4/3k^2$. Substituting this expression into Eq. (39) we obtain

$$Q^2=4\left[\overline{\lambda}+\left(\frac{1}{3}\frac{\alpha^4}{k^2}-1\right)\overline{\lambda^2}\right]. \tag{40}$$

Differentiating the latter equation with respect to $\bar{\lambda}$ and setting the derivative equal to zero, we find the critical value

$$\bar{\lambda}_c = \frac{1}{2-2\alpha^4/3k^2} = \frac{3}{4}\,\frac{\alpha^2-1}{\alpha^2-3/2},$$

which corresponds to the maximum

$$Q^2(\alpha^2) = \frac{3}{2}\,\frac{\alpha^2-1}{\alpha^2-3/2}. \tag{41}$$

In particular, when $\alpha^2=2$ (which corresponds to a flat rotation curve), $Q_c^2=3$. This maximum is achieved when $g^2=16/27$ and $\lambda=1.5$, i.e., when $m=2v_0/3c_s$. A plot of $Q_c^2(\alpha^2)$ is shown in Fig. 5c.

We note that repeated attempts to obtain a general stability criterion for disks with a two-dimensional pressure have previously been undertaken.[19–21] They were made, however, within the approximation of tightly wound spirals ($s=k_\varphi/k_r \ll 1$,) which has traditionally been used in all such investigations. The inadequacy of such an approach can be seen from the results of the theory itself. The stability criterion obtained in it can be written, for example, in the form

$$c_s^2 > c_0^2[1-2s^2 r(\Omega^2)'/\kappa^2], \tag{42}$$

where $c_0=\text{const}$. It is seen from this equation that, as s^2 increases, larger values of c_s are required to suppress the stability. In this situation it is clear that the answer to the main question in the theory, i.e., the conditions for the existence of a completely stable disk, is beyond its scope. Of course, it can be expected *a priori* that perturbations with $s \gg 1$ ($k_\varphi \gg k_r$) are stable, and that perturbations with $s \approx 1$ are highly unstable.

Equation (39) can be compared with the analogous equation of Lin and Lau,[19] which was also derived in the approximation of tightly wound spirals, $g^2 \ll 1$:

$$Q^2 = 4[\bar{\lambda}-\bar{\lambda}^2/(1+g^2\bar{\lambda}^2)]. \tag{43}$$

Although this equation was derived for $g^2 \ll 1$, Lin and Lau used it for $g^2 \approx 1$. As expected, Eq. (39) and the equation derived from (43) for small $g^2\bar{\lambda}^2 \ll 1$ coincide, but they do not have anything in common when $g^2\bar{\lambda}^2 \approx 1$. Plots of $Q^2(\bar{\lambda})$ borrowed from Lin and Lau's paper[19] for several values of g are presented in Fig. 5a. It is seen that the $Q_g^2(\bar{\lambda})$ curves described by Eq. (43) are monotonic and, for this reason alone, are incorrect. It should also be borne in mind that the evolution of each perturbation at any moment does, in fact, take place in accordance with the instantaneous value of the radial wave number q, which varies linearly with time. The time when the region of unstable values of q is intersected, as well as the degree of enhancement, are constrained (and are small near $Q_g=Q_c$). When $Q_g > Q_c$, the enhancement region vanishes entirely; in this case ''complete stability'' of a disk in the sense that perturbations do not grow (at least not quasiexponentially) even during some constrained time interval can be guaranteed.

3. STELLAR DISKS

The problem of the evolution of a perturbation assigned at a certain initial time $t=t_0$ was previously solved in a paper by Julian and Toomre,[15] who used the same basic approximations as we shall use below (a local Cartesian coordinate frame, a linear law for variation of the angular velocity, etc.). Several very interesting results were obtained in the paper just cited. Among them we can single out the conclusion that any perturbations in differentially rotating stellar disks that satisfy Toomre's criterion are asymptotically stable. However, under the approach adopted in Julian and Toomre's work it was not possible to obtain a useful tool for investigating stability, such as a local dispersion relation (this is accomplished in the present paper).

As in Sec. 2.2, we use the model of an infinitely thin disk with a surface density $\sigma = \int f dv_r dv_\varphi$, where $f(r,\varphi,v_r,v_\varphi;t)$ is the distribution function of stars in the (r,φ) cylindrical coordinate system, and v_r and v_φ are the components of the velocity of a star in a reference frame rotating with the angular velocity $\Omega(r)$. The function f obeys the equation

$$\frac{\partial f}{\partial t} + v_r\frac{\partial f}{\partial r} + \left(\Omega + \frac{v_\varphi}{r}\right)\frac{\partial f}{\partial \varphi} + \left(-\frac{\partial \Phi}{\partial r} + \Omega^2 r + 2\Omega v_\varphi + \frac{v_\varphi^2}{r}\right)\frac{\partial f}{\partial v_r} - \left(\frac{1}{r}\frac{\partial \Phi}{\partial \varphi} + \frac{\kappa^2}{2\Omega}v_r + \frac{v_r v_\varphi}{r}\right)\frac{\partial f}{\partial v_\varphi} = 0. \tag{44}$$

Assuming that the equilibrium distribution function $f_0 = A\exp(-Q_0)$ ($A=\text{const}$, $Q_0 = v_r^2/2c_r^2 + v_\varphi^2/2c_\varphi^2$), we linearize (44) by plugging in $\Phi = \Phi_0 + \Phi_1$ ($|\Phi_1| \ll |\Phi_0|$) and $f = f_0(1+\psi)$ ($|\psi| \ll 1$) and going over to the dimensionless velocities ξ and η, where $v_r = \xi V_1$ and $v_\varphi = \eta V_2$ [$V_1 = 2\Omega V_0/\kappa$, $V_2 = V_0$, $V_0 = \Omega(r)r$]:

$$-i\omega_*\psi + im\Omega\eta\psi + 2\frac{\Omega_0^2}{\kappa}r\xi\frac{\partial\psi}{\partial r} + \kappa\eta\frac{\partial\psi}{\partial\xi} + \frac{\kappa}{2\Omega^2 r}\frac{\partial\Phi_1}{\partial r}\frac{\partial Q_0}{\partial\xi} - \kappa\eta\frac{\partial\psi}{\partial\eta} + \frac{im}{\Omega r^2}\Phi_1\frac{\partial Q_0}{\partial\eta} = 0. \tag{45}$$

Here only the first-order terms with respect to the small quantities Δ/r_0 and c_s/V_0 have been left and the equilibrium condition $\partial\Phi_0/\partial r = \Omega_r^2$ has been taken into account. Here the perturbations are assumed to be proportional to $\exp(-i\omega t + im\varphi)$.

Going over to the local Cartesian coordinate frame near the corotation radius $r=r_0$ [$\Omega(r_0)=\Omega_p$, $x=(r-r_0)/\Delta$, $\Delta = \kappa/m|\Omega_0'|$], introducing the polar coordinates τ and s in the space of the velocities $\xi = \tau\cos s$ and $\eta = \tau\sin s$, and using the notation in Sec. 2.2, we obtain the following equation for ψ:

$$\frac{\partial\psi}{\partial s} + iz\psi - \frac{\tau}{\delta}\left[i\alpha^2\left(\frac{1}{2}\psi + \phi\frac{\partial Q_0}{\partial\tau^2}\right)\sin s + \left(\frac{1}{2}\frac{\alpha^4}{k^2}\frac{\partial\psi}{\partial x} + \frac{d\phi}{dx}\frac{\partial Q_0}{\partial\tau^2}\right)\cos s\right] = 0, \tag{46}$$

where we have introduced the notation $\delta=\Delta/r_0$ and $\phi=\Phi_1/V_0^2$. In (46) we go over to the Fourier transforms $[\psi_q=\int_{-\infty}^{\infty}dx\psi(x)e^{-iqx}]$; in the new variables p and q' ($q=q'$ and $s=p-q'$) Eq. (46) becomes

$$\frac{\partial\psi_q}{\partial q}+a_q\psi_q=-b(q), \tag{47}$$

where

$$a(q)=\nu+i\frac{\tau\alpha^2}{2\delta}[\sin(p-q)+q\alpha^2\cos(p-q)/k^2],$$

$$b(q)=i\frac{\tau}{\delta}\phi_q\frac{\partial Q_0}{\partial\tau^2}[\alpha^2\sin(p-q)+q\cos(p-q)].$$

The solution of Eq. (47), which satisfies the condition $\psi_q\to 0$ as $q\to\infty$, is

$$\psi_q=\exp[-F(q,p,\tau)]\int_q^{\infty}b(q')\exp[F(q',p,\tau)]dq', \tag{48}$$

where

$$F(q,p,\tau)=\nu q+i\frac{\tau\alpha^4}{2\delta k^2}[\alpha^2\cos(p-q)-q\sin(p-q)]. \tag{49}$$

We introduce the variables $\xi=\tau\cos p$ and $\eta=\tau\sin p$. Clearly, $d\xi d\eta=\tau d\tau dp=\tau d\tau ds$ for a fixed q. In these variables

$$F(q,\xi,\eta)=\nu q+i\frac{\alpha^4}{2\delta k^2}(\xi f_\xi+\eta f_\eta),$$

$$f_\xi=\alpha^2\cos q+q\sin q,\quad f_\eta=\alpha^2\sin q-q\cos q,$$

$$b(q,\xi,\eta)=i\frac{\varphi_q}{2\delta}\left(\frac{\partial Q_0}{\partial\eta}f_\xi-\frac{\partial Q_0}{\partial\xi}f_\eta\right).$$

After some elementary calculations, the surface density $\sigma_q=V_1V_2\int_{-\infty}^{\infty}d\xi\int_{-\infty}^{\infty}d\eta f_0\psi_q$ is obtained in the form

$$\sigma_q=\frac{\sigma_0\alpha^4}{4\delta^2k^2}\int_q^{\infty}dq'\phi_{q'}h(q,q')\exp[\nu(q'-q)-xg(q,q')], \tag{50}$$

where $x=c_r^2/2\Delta^2\kappa^2$, and

$$h(q,q')=f_\xi(q')f_\eta(q)-f_\eta(q')f_\xi(q),$$

$$g(q,q')=[f_\xi(q')-f_\xi(q)]^2+[f_\eta(q')-f_\eta(q)]^2.$$

The functions h, g, and σ_q can be brought into the form

$$h(q,q')=(qq'+\alpha^4)\sin(q-q')-\alpha^2(q-q')\times\cos(q-q'), \tag{51}$$

$$g(q,q')=q^2+q'^2+2\alpha^4-2[(qq'+\alpha^4)\times\cos(q-q')+\alpha^2((q-q')\sin(q-q')], \tag{52}$$

$$\sigma_q=\frac{\sigma_0}{\kappa^2\Delta^2}\int_q^{\infty}dq'\Phi_{q'}h(q,q')\times\exp[\nu(q'-q)-xg(q,q')]. \tag{53}$$

The last equality specifies the density response to a potential perturbation. Another relation between σ_q and Φ_q is obtained from Poisson's equation:

$$\sigma_q=-\frac{\Phi_q\sqrt{q^2+k^2}}{2\pi G\Delta}. \tag{54}$$

Introducing the notation $c=\kappa^2\Delta/\pi G\sigma_0$ and equating the left-hand sides of Eqs. (53) and (54), we obtain an integral equation of the Volterra type for the Fourier amplitude of the potential

$$\frac{c}{2}\Phi_q\sqrt{q^2+k^2}=-\int_q^{\infty}dq'\Phi_{q'}K(q,q')\exp[\nu(q'-q)], \tag{55}$$

where the kernel is defined by the formula

$$K(q,q')=h(q,q')\exp[-xg(q,q')]. \tag{56}$$

It follows from the obvious properties of the kernel (56), viz., $K(q',q)=-K(q,q')$ and $K(-q,-q')=-K(q,q')$, that Eq. (55) has either even or odd solutions.

It can be shown for $x=0$ that the integral equation (55) reduces to a differential equation that coincides with Eq. (34) when $a=0$, i.e., in the case of a cold gaseous disk. For $x\neq 0$ the integral equation (55) cannot be reduced to a differential equation.

Replacing q by $-k_\varphi\Delta\tau$, as in (9) or (36), we obtain the evolution equation for $\Phi_\tau\equiv\Phi_q$ ($q=-k_\varphi\Delta\tau$):

$$\frac{c}{2}\sqrt{1+\tau^2}\Phi_\tau+\int_{-\infty}^{\tau}d\tau'\Phi_{\tau'}\overline{K(\tau,\tau')}=0, \tag{57}$$

where we have introduced the notation $\overline{K(\tau,\tau')}=K(q=-k_\varphi\Delta\tau,q'=-k_\varphi\Delta\tau')$ and, for brevity, have omitted the term containing the small parameter ν in the exponential function. To derive a stability criterion, we consider solutions of Eq. (57) of the form $\exp(-i\beta\tau)$, and we then obtain

$$\frac{c}{2}\sqrt{1+\tau^2}+\int_0^{\infty}dt\,e^{-i\beta t}K(\tau,\tau-t)=0, \tag{58}$$

or, going back from τ to q, we obtain

$$\frac{c}{2}\sqrt{q^2+k^2}+\int_0^{\infty}dt\,e^{-itz}\overline{K(t,q)}=0, \tag{59}$$

where $\overline{K(t,q)}=K(q',q)$, and $q-q'=t$. This relation is the dispersion relation which we are seeking, in which $z=\beta/k_\varphi\Delta$,

$$h(t,q)=\alpha^2t\cos t-(q^2-qt+\alpha^4)\sin t,$$

$$g(t,q)=t^2+2(q^2-qt+\alpha^4)(1-\cos t)-2\alpha^2t\sin t.$$

The solution of Eq. (58) gives the local (instantaneous) value of the frequency β, and the integral (true) dependence of Φ_τ can be represented qualitatively in the form[22] $\Phi_\tau\approx\exp[-i\int_{\tau_0}^{\tau}\beta(\tau')d\tau']$. We note that $x=Q_0^2g^2/4\alpha^2$, where

now $Q_0=\kappa c_r/2\pi G\sigma_0$. For convenience in making a comparison with the results in Sec. 2.2, we introduce the notation $\bar{\lambda}=\lambda/\lambda_c$ and $u^2=(1+2/g^2\bar{\lambda}^2)/\alpha^2$. Then $q=\alpha^2\sqrt{u^2-1}$, and $q^2+\alpha^4=\alpha^4u^2$. Using these relations, we rewrite the dispersion relation (59) in the final form

$$1+\bar{\lambda}g^2R_z=0, \tag{60}$$

where

$$R_z=\int_0^\infty dt\ e^{-itz}H(u^2,t),$$

$$H(u^2,t)=\left[\frac{1}{2}t\cos t+\left(\frac{1}{2}t\sqrt{u^2-1}-\frac{\alpha^2}{2}u^2\right)\sin t\right]\times\exp\left\{-Q_0^2g^2\left[\frac{t^2}{4\alpha^2}+\left(\frac{\alpha^2}{2}u^2-\frac{1}{2}t\sqrt{u^2-1}\right)\times(1-\cos t)-\frac{1}{2}t\sin t\right]\right\}.$$

When $z=0$, from Eq. (60) we find the sought-for criterion of marginal stability [the $Q_0^2(\bar{\lambda})$ curve]:

$$f(\bar{\lambda},Q_0^2)=1+\bar{\lambda}g^2R_0=0,$$
$$R_0=\int_0^\infty dtH(u^2,t). \tag{61}$$

It follows from the condition $u^2>1$ (q is then real) that $0<\bar{\lambda}\leqslant\sqrt{2/g^2(\alpha^2-1)}\equiv\bar{\lambda}_g$, and that the marginal stability curve $Q_0^2(\bar{\lambda})$ lies in the range $(0,\bar{\lambda}_g)$.

We shall show that in the limit $g\to0$ Eq. (61) coincides (as it should) with the marginal stability criterion derived by Toomre[7] for radial perturbations of a stellar disk. We introduce the notation $z\equiv Q_0^2/\bar{\lambda}^2$. We can then write Toomre's marginal stability criterion[7] for axisymmetric disk perturbations in the form

$$\bar{\lambda}z=1-e^{-z}I_0(z), \tag{62}$$

where $I_0(z)$ is a modified Bessel function. For small g

$$H(u^2,t)\approx-\frac{\alpha^2}{2}u^2\sin t\times\exp\left[-Q_0^2g^2\left(\frac{t^2}{4\alpha^2}+\frac{\alpha^2u^2(1-\cos t)}{2}\right)\right].$$

Since here $\alpha^2g^2u^2/2\approx\bar{\lambda}^{-2}$, the stability criterion (61) can be written for small g in the form

$$1-\frac{e^{-z}}{\bar{\lambda}z}\int_0^\infty dt\sin t\exp(-\varepsilon t^2+z\cos t)=0, \tag{63}$$

where we have introduced the notation $\varepsilon=Q_0^2g^2/4\alpha^2$ ($\varepsilon\to0$). Integrating (63) in parts, we obtain

$$\bar{\lambda}z=1-\exp(-z)R_\varepsilon,$$

where

$$R_\varepsilon=2\varepsilon\int_0^\infty dt\ t\exp(-\varepsilon t^2+z\cos t). \tag{64}$$

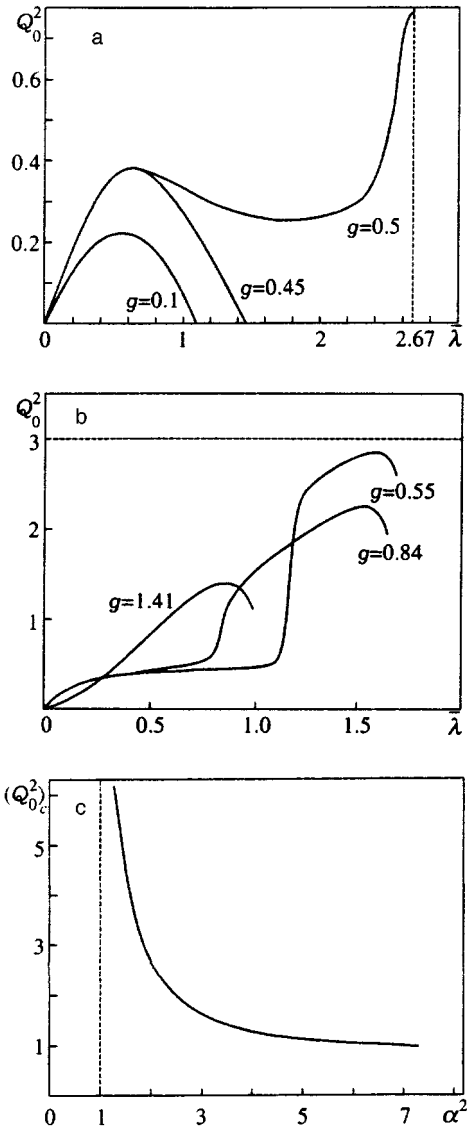

FIG. 6. Plots of the marginal stability $Q_0^2(\bar{\lambda})$ for a stellar disk with a flat rotation curve ($\alpha^2=2$) and $g=0.1$, 0.45, 0.50 (a) or $g=0.55$, 0.84, 1.414 (b), as well as the dependence of the critical value $(Q_0^2)_c(\alpha^2)$ for a stellar disk (c).

Since $R_0=I_0(z)$, we do, in fact, obtain Toomre's stability criterion (62) for $\varepsilon\to0$.

The solutions of Eq. (61) for a flat rotation curve ($\alpha^2=2$) were found numerically for different g (Figs. 6a and 6b). It is seen that a critical value $(Q_0^2)_c$ is also obtained here for a certain g, by analogy with the case of a gaseous disk. Figure 6c presents a plot of $(Q_0^2)_c(\alpha^2)$. If the disk parameters α^2 and Q_0^2 are represented by a point in the (α^2,Q_0^2) plane lying above the $(Q_0^2)_c(\alpha^2)$ curve, the corresponding model will be stable toward all small perturbations in the plane of the disk. We note that for a disk rotating as a solid body ($\alpha^2\to\infty$) the critical value $(Q_0^2)\to0$ (in a gaseous disk the corresponding $Q_c^2\to3/2$). In addition, here $(Q_0^2)_c\to\infty$ as $\alpha^2\to1$ (while in a gaseous disk $Q_c^2\to\infty$ as $\alpha^2\to3/2$).

The relationship between Q_0 and Toomre's instability parameter $Q=\kappa c_r/3.36G\sigma_0$ (which is equal to unity at the stability boundary of radial perturbations) has the form

$Q=2\pi Q_0/3.36$; accordingly, for a flat rotation curve ($\alpha^2=2$), where $(Q_0)_c^2\approx 2.85$, we have $Q_c\approx 3.15$. The latter value is consistent with the results of Toomre's[23] numerical (N-body) experiments with stellar disks: according to his data, the disks became completely stable specifically when $Q\geqslant 3$.

It is striking that the values of Q_c are significantly higher for stellar disks than for gaseous disks. For example, in the case of a flat rotation curve ($\alpha^2=2$), which is most important for galaxies, these values of Q_c are equal, respectively, to $\sqrt{3}$ and ≈ 3. We mention one, probably the main, cause of the great instability of nonaxially symmetric perturbations in stellar systems. It is that a differentially rotating stellar disk, in contrast with a gaseous disk, which is always isotropic, is colder in the azimuthal direction and is, therefore, more unstable in the ''Jeans'' sense. This follows from the known relation (which is attributed to Lindblad[24]) between velocity dispersions in the radial (c_r) and azimuthal (c_φ) directions:

$$c_r=\frac{2\Omega(r)}{\kappa(r)}c_\varphi>1 \quad (\Omega'<0). \tag{65}$$

The simplest way to prove the correctness of relation (65) is as follows. The distribution function f_0 of stars in an equilibrium disk must satisfy the kinetic equation (in cylindrical coordinates)

$$v_r\frac{\partial f_0}{\partial r}-\frac{\partial\Phi_0}{\partial r}\frac{\partial f_0}{\partial v_r}+\frac{v_\varphi^2}{r}\frac{\partial f_0}{\partial v_r}-\frac{v_r v_\varphi}{r}\frac{\partial f_0}{\partial v_\varphi}=0, \tag{66}$$

where Φ_0 is the equilibrium gravitational potential. The distribution function of a cold disk f_0 is proportional to $\delta(v_r)\delta(v_\varphi-\Omega r)$, where $r\Omega^2=d\Phi_0/dr$. Let us now represent the distribution function of a disk system with nearly circular stellar orbits in the form of a formal series in δ functions and their derivatives with respect to the arguments v_r and $v_\varphi-\Omega r$:[2,3]

$$\begin{aligned} f_0=&a_1\delta(v_r)\delta(v_\varphi-\Omega r)+b_1\delta'(v_r)\delta(v_\varphi-\Omega r)\\ &+b_2\delta(v_r)\delta'(v_\varphi-\Omega r)+c_1\delta''(v_r)\delta(v_\varphi-\Omega r)\\ &+c_2\delta'(v_r)\delta'(v_\varphi-\Omega r)+c_3\delta(v_r)\delta''(v_\varphi-\Omega r). \end{aligned} \tag{67}$$

Substituting (67) into (66) and setting the coefficients in various combinations of derivatives of the δ functions equal to zero, we obtain a system of equations, one of which is $c_3=\kappa^2 c_1/4\Omega^2$, which is equivalent to (65), since, clearly, $c_1=\sigma_0 c_r^2/2$ and $c_3=\sigma_0 c_\varphi^2/2$ (σ_0 is the surface density at the point r). We note that the inequality (65) is analogous (and apparently opposite) to the equation presented at the beginning of this paper for the aspect ratio of an epicycle [Eq. (2)].[7)] We also note that the argument that the Lindblad relation between the velocity dispersions can be the cause of the great instability of the ''oblique'' modes in differentially rotating collisionless disks was previously advanced by Hunter[25] (however, no general stability criteria were previously derived).

4. CONCLUSIONS

We have thus obtained the criteria for the local stability of perturbations in gaseous and stellar disks. These criteria are conveniently formulated in terms of Toomre's stability parameter, which, by definition, is equal to unity on the stability boundary of radial perturbations. It has been found that the stabilization of arbitrary perturbations requires a value of Q greater than a certain critical value Q_c, and that Q_c is significantly greater than unity; i.e., it can be stated that perturbations which destroy the original axial symmetry of a disk are more unstable than radial perturbations. In addition, the values of Q_c for stellar disks are appreciably greater than the values for gaseous disks. One, probably the main, cause of the great instability of stellar systems was considered at the end of the preceding section.

The earlier attempts[19–21] were unsuccessful, because the authors restricted their analysis to the approximation of tightly wound spirals ($k_r\gg k_\varphi$). The attempts to extend the results thus obtained to perturbations with $k_r\approx k_\varphi$ and $g^2\approx 1$ were futile: they led to conclusions which are qualitatively incorrect (compare Figs. 5a and 5b). The interpretations of the local dispersion equation, as well as the marginal stability conditions that we gave at the end of Sec. 2.2, of course, differ from the usual interpretation. We obtained sufficient stability criteria, which guarantee the absence of any (even a small) time interval of quasiexponential perturbation growth. These stability criteria are clearly more than sufficient, since perturbations scarcely grow near $Q=Q_c$. At the same time, under the usual interpretation of the marginal curve as the boundary of exponential instability, it is assumed that unstable perturbations can grow at an unlimited rate after a sufficiently long time even near the stability boundary.

Returning to Fig. 4, which presents the solutions of the evolution equation (37), we note that, in general, a perturbation can grow for two reasons: 1) due to the presence of intervals (τ_1,τ_2) of quasiexponential growth, where the square of the characteristic frequency is negative, i.e., $\omega^2=\beta<0$, and 2) due to a decrease in β (at $\tau<0$). If we use an estimate that is rigorous when the adiabatic invariant is conserved ($I=E/\omega=\text{const}$), the variation of the amplitude a due to the second factor follows the law $a\propto 1/\sqrt{\omega}$. This mechanism remains the only one in cases in which there is no interval with quasiexponential growth (as in Fig. 4c). We note, however, that a significant increase in the perturbation amplitude occurs only if there is a quasiexponential interval.

At first glance, it appears that a determination of the conditions under which perturbations are enhanced by a certain factor would be more appropriate in the situation described. It is natural to proceed in this way in cases where the level of the initial perturbations is known (it is most often the level of thermal fluctuations). There is no such certainty in the problems which interest us: in real gravitating systems perturbations which greatly surpass the thermal fluctuations can appear for many reasons (for example, as a result of tides).

Of course, global modes can be analyzed only by numerical methods of a different kind or (analytically) in the case of a very small number of ultrasimplified models like

solid-body disk rotation.[2,3] However, having simple local stability criteria, we can also confidently predict the properties of global modes. Unfortunately, the only disks that have heretofore been systematically studied (disks with uniform rotation) are least suitable for comparison with our local analysis (for example, the characteristic scale that we adopted $\Delta \to \infty$ when $\Omega' \to 0$).

In conclusion, we note that in our study we restricted the analysis to a treatment of perturbations which are symmetric with respect to the $z=0$ equatorial plane of the disk (which do not cause it to bend). These perturbations are associated with such phenomena as, for example, the appearance of the spiral structure of galaxies. Nevertheless, a flexural instability, which is similar to the hose instability in hydrodynamics and plasmas,[26] can develop in stellar disks. It also plays an important role. For example, this instability imposes very significant constraints on the maximum possible degree of flattening of equilibrium systems, such as elliptical galaxies. The theory of hose instability in gravitating systems was presented in detail in Refs. 2 and 3.

We thank the referee for several important comments that allowed us to improve the quality of this paper.

This work was supported, in part, by the Russian Fund for Fundamental Research (Grants Nos. 96-02-17792 and 96-02-18239).

[1)]In principle, for collisionless systems, such as strongly flattened clusters of galaxies, elliptic galaxies, or the central regions of the stellar components of disk galaxies, this is not the only possibility: in this case the disk shape can be partially (or even completely) attributed to the anisotropy of the pressures in the rotation plane and in the direction perpendicular to it. However, most astrophysical disks are highly flattened mainly as a result of rapid rotation. In this paper we restrict the analysis to such disks (for more details on the alternative possibilities see Refs. 2 and 3).

[2)]Actually, Toomre[7] defined Q somewhat differently: $Q=\kappa c_r/3.36G\sigma_0$. Just this quantity is equal to unity on the stability boundary of radial perturbations of stellar disks, which were considered by Toomre (c_r is the dispersion of the radial velocities of the stars). For further details, see Sec. 3.

[3)]More precisely, this refers to perturbations which maintain a generalized vortex $(\mathrm{curl}_z\ \mathbf{v}/\sigma)_1=0$.

[4)]As is widely known, the characteristic solutions of the Rayleigh solution can exist provided $v_0''=0$ anywhere within the flow region. In the case of a rotating flow of a gravitating compressible gas considered by us, a similar condition may be required for the existence of characteristic solutions.

[5)]We note that the approach of Goldreich and Lynden-Bell utilizes several assumptions, which unavoidably render the results obtained approximate (we, however, believe that the exact results, which require a significantly more complicated derivation, do not differ strongly from the results obtained by Goldreich and Lynden-Bell[14] and by us in this paper). First of all, we are dealing with the use of an approximation of the vertical equilibrium: it is assumed that the dependence of the perturbed density on z is the same as for the unperturbed density. The exact mechanisms for the establishment of equilibrium along z in real astrophysical disks should be taken into account by a detailed theory in the future. A discussion of this problem (as well as its formulation) can be found in the work of Fridman and Khoruzhiĭ.[16] Another deficiency of Goldreich and Lynden-Bell's theory is the use of the formulas for an isolated disk to describe the equilibrium along z: this actually means that the external halo is assumed to be cylindrical (while it would be more natural to assume that it is spheroidal).

[6)]Thus, our approximation is formally not applicable when $m \approx 1$ (which is usually the case for global galactic spiral modes).

[7)]The contradiction between Eqs. (65) and (2) is only apparent. It is due to the fact that in the former case the perturbed azimuthal velocity is measured relative to $r\Omega(r)$ (which has a different value at each point r), and in the latter case it is measured relative to $r_0\Omega(r_0)$, where r_0 is the position of the center of the epicycle.

[1] V. L. Polyachenko, Zh. Éksp. Teor. Fiz. **101**, 1409 (1992) [Sov. Phys. JETP **74**, 755 (1992)].
[2] V. L. Polyachenko and A. M. Fridman, *Equilibrium and Stability of Gravitating Systems* [in Russian], Nauka, Moscow (1976).
[3] A. M. Fridman and V. L. Polyachenko, *Physics of Gravitating Systems*, Springer-Verlag, New York (1984), Vols. 1 and 2 [Russ. original, Nauka, Moscow (1976)].
[4] W. L. W. Sargent, P. J. Young, A. Boksenberg, K. Shortridge, C. R. Lynds, and F. D. A. Hartwick, Astrophys. J. **221**, 731 (1978).
[5] M. J. Duncan and J. C. Wheeler, Preprint, Department of Astronomy, University of Texas, Austin (1979).
[6] A. J. Newton and J. J. Binney, Mon. Not. R. Astron. Soc. **210**, 711 (1984).
[7] A. Toomre, Astrophys. J. **139**, 1217 (1984).
[8] Ya. B. Zel'dovich and I. D. Novikov, *Structure and Evolution of the Universe* [in Russian], Nauka, Moscow (1975).
[9] R. W. Miller, K. H. Prendergast, and W. J. Quirk, Astrophys. J. **161**, 903 (1970).
[10] F. Hohl, Astrophys. J. **168**, 343 (1971).
[11] C.-C. Lin, *The Theory of Hydrodynamic Stability*, Cambridge University Press, Cambridge (1966) [Russ. transl., IL, Moscow (1958)].
[12] K. M. Case, Phys. Fluids **3**, 149 (1960).
[13] D. G. Lominadze, G. D. Chagelashvili, and R. G. Chanashvili, Pis'ma Astron. Zh. **14**, 856 (1988) [Sov. Astron. Lett. **14**, 364 (1988)].
[14] P. Goldreich and D. Lynden-Bell, Mon. Not. R. Astron. Soc. **130**, 97,125 (1965).
[15] W. H. Julian and A. Toomre, Astrophys. J. **146**, 810 (1966).
[16] A. M. Fridman and O. V. Khoruzhiĭ, N. N. Gor'kavyĭ and A. M. Fridman, *Physics of Planetary Rings* [in Russian], Nauka, Moscow (1994).
[17] V. L. Polyachenko, S. M. Churilov, and I. G. Shukhman, Astron. Zh. **57**, 497 (1980) [Sov. Astron. **24**, 287 (1980)].
[18] D. Lynden-Bell and A. J. Kalnajs, Mon. Not. R. Astron. Soc. **157**, 1 (1972).
[19] C. C. Lin and Y. Y. Lau, Stud. Appl. Math. **60**, 97 (1979).
[20] G. Bertin and C. C. Lin, in *Evolution of Galaxies (Proceedings of the 10th IAU Regional Astronomy Meeting)*, J. Palous (ed.), Publications of the Astronomy Institute, Czechoslovak Academy of Sciences (1988), p. 255.
[21] A. G. Morozov, Astron. Zh. **57**, 805 (1980) [Sov. Astron. **24**, 391 (1980)].
[22] G. S. Bisnovatyĭ-Kogan and Ya. B. Zel'dovich, Astron. Zh. **47**, 942 (1970) [Sov. Astron. **14**, 758 (1971)].
[23] A. Toomre, in *Structure and Evolution in Normal Galaxies*, S. M. Fall and D. Lynden-Bell (eds.), Cambridge University Press, Cambridge (1981), p. 111.
[24] B. Lindblad, Stockholms Obs. Ann. **20**, (6) (1958).
[25] C. Hunter, Astrophys. J. **181**, 685 (1973).
[26] A. B. Mikhaĭlovskiĭ, *Theory of Plasma Instabilities*, Consultants Bureau, New York (1974).

Translated by P. Shelnitz
Edited by S. J. Amoretty

Spin state tomography

V. I. Man'ko and O. V. Man'ko

P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 18 February 1997)
Zh. Éksp. Teor. Fiz. **112**, 796–804 (September 1997)

A scheme for measuring the quantum state for an arbitrary spin is proposed that is analogous to the symplectic tomography scheme used to measure quantum states associated with continuous observables such as position and momentum. An invariant form for the spin state density operator is derived in terms of an integral, over the angles which specify the quantization axis, of a product of the measured probability of the values of the spin along a chosen direction and spherical harmonics summed with Clebsch–Gordan functions. © *1997 American Institute of Physics.* [S1063-7761(97)00209-6]

1. INTRODUCTION

In quantum mechanics there is a radical change in the concept of the ''state of a system'' compared to classical mechanics.[1] The state of a quantum mechanical system is described either by a wave function[2] (for pure states) or by a density matrix[3,4] (for mixed states). From the very beginning of the development of quantum mechanics, attempts have been made to explain[5–7] it ''classically,'' but without success. The Wigner function,[8] as well as the Glauber–Sudarshan P-distribution[9,10] and the Husimi–Kano Q-function,[11,12] have been introduced in attempts to bring the description of quantum states in terms of the language of the density matrix closer to a classical description that involves a comparison to the state of a positive normalized probability distribution function of the measured quantity (for example, position or momentum). All of these functions depend on the variables q and p and are normalized, and for individual quantum states (e.g., of an oscillator at temperature T) they are similar to the classical distribution functions in phase space. Obviously, however, none of them is a probability distribution, since the uncertainty relations for position and momentum do not allow simultaneous measurements of these conjugate variables, so that for a quantum mechanical system (e.g., an oscillator), no phase space distribution function exists. For this reason, the Wigner, Glauber–Sudarshan, and Husimi–Kano functions, which specify a fully mixed quantum state and are related to one another and to the density matrix in the position representation by invertable integral transforms, have been referred to as quasidistributions.

Quasidistributions differ significantly from probability distributions. In particular, the Wigner function, which is a real function, can take negative values, so it cannot be a distribution function, which can only take nonnegative values. The Husimi–Kano function is normalized and only takes nonnegative values, but its arguments q and p are not a simultaneously measurable position and momentum; thus, this function is also a quasidistribution. Nevertheless, it turns out that a quantum state of a system can be described not only by quasiprobabilities, but also by legitimate probability distribution functions.

The measurable quantities in quantum mechanics are those such as position, momentum, and energy. But the following physical statement of the problem is also possible: how can the quantum state itself of a system (i.e., the Wigner function, density matrix) be measured? In solving this problem, it has been possible to show that the state of a quantum mechanical system can be specified not only by a quasidistribution such as the Wigner function, but also by actual probability distributions. These distributions are referred to as marginal distributions, since besides depending on a random physical quantity, they depend on additional parameters that describe different reference frames in the classical phase space of the system.

The problem of measuring quantum states has been widely studied of late, both theoretically and experimentally. Thus, a relationship was found[13,14] between the Wigner function[8] and the measured marginal probability distribution for a homodyne observable, which is a coordinate that has been rotated by a specified angle in the phase space of the system. The Wigner function of a one-dimensional system is expressed in terms of this measurable normalized positive distribution function using the Radon transform (with integration over the rotation angle in phase space) employed in conventional medical tomography. For this reason, the scheme for measuring the quantum state for a continuous observable such as position or momentum is referred to as an optical tomography scheme, and this scheme has been used in experiments to reconstruct the quantum state of an electromagnetic wave mode[15] and in molecular spectroscopy.[16]

The optical tomography scheme was modified[17,18] to a symplectic tomography scheme that uses a measurement of the normalized and positive distribution function for a continuous observable that is the position measured, not in one frame of reference in phase space, but in an ensemble of frames of reference related to one another by linear transformations involving rotations and changes of scale, in order to reconstruct the quantum state. In this scheme, referred to as symplectic tomography, the Wigner function is recovered from the experimental data using a Fourier transform, rather than a Radon transform, and this made it possible to generalize symplectic tomography to multimode systems.[18]

The optical tomography scheme was formulated[19] in an invariant form that is independent of the quantum mechanical representation employed. An expression for the density

 1063-7761/97/090430-05$10.00

operator (and not just for the Wigner function) is obtained in terms of the measured probability distribution. An invariant expression for the density operator in symplectic tomography[18] is used[20] to obtain a ''classical'' equation (equivalent to the Schrödinger equation for the wave function) for the temporal evolution of a quantum state described by a positive marginal distribution (of a continuous observable).

One more modification of the tomographic method for measuring the quantum state has been examined.[21–23] The proposed scheme involves measuring the distribution over a discrete number of quanta (photons) in the mode being considered, which depends further on the controlled phase and amplitude of an external classical field superimposed on a signal field that is in a quantum state. The reconstruction of even and odd coherent states of ions[24] in a trap[25] has been discussed.[26–30] States of this type have been realized experimentally for ions in traps[31] and for modes of the electromagnetic field in a microcavity.[32]

Spin, on the other hand, is a purely quantum mechanical entity with no classical limit. Its states are usually described by spinors. The following physical question is also important for this specifically quantum mechanical quantity: how can we measure the quantum state of a spin degree of freedom? An attempt has been made to introduce a description of the states of a discrete quantum mechanical observable such as spin using a classical distribution.[33] A scheme for coupling the spin density matrix with an observable marginal distribution for an arbitrary spin has been discussed briefly[34] by analogy with papers on optical and symplectic tomography in the noninvariant form,[13,14,17] and using a group theoretical treatment of the tomographic schemes.[23,25]

The goal of the present paper is to derive in detail and analyze a method for the tomography of spin states that was formulated briefly in an earlier paper,[34] and to obtain an invariant expression for the density of states operator of an arbitrary spin j in terms of the measured probability of the projection of the spin in any direction. This density operator will be expressed in terms of an integral over solid angle of the measured distribution, which depends further on the two angles that specify this direction, and a standard operator, which depends on spherical harmonics, and therefore on these angles. Although this expression is analogous to the Radon transform for optical tomography, it differs substantially from it. Thus, in tomographic methods for measuring a quantum state that employ invertable integral transforms of the density matrix and of the measurable probability distribution function (marginal distribution), it is assumed implicitly that the state of the quantum system can be specified, not by the density matrix, but by the probability distribution function of the measurable quantity. This definition of a quantum state is discussed explicitly elsewhere.[36,37] Nevertheless, the discussion of the tomography of quantum states has been associated with measurements of quantities like position, which have a classical limit.

2. DENSITY MATRIX AND MARGINAL DISTRIBUTION

In this section we introduce a family of positive normalized probability distributions for the projection of the spin, and show that the density matrix of an arbitrary spin state can be expressed in terms of it.

Let us consider a state, for arbitrary spin j, which is described by a hermitian density matrix with matrix elements $\rho^{(j)}_{mm'}$ corresponding to the density operator $\hat{\rho}(j)$ in a basis of eigenvectors $|jm\rangle$ for the projection of the spin and the square of the spin along the z axis, i.e.,

$$\hat{j}_3|jm\rangle = m|jm\rangle, \quad \hat{j}^2|jm\rangle = j(j+1)|jm\rangle, \tag{1}$$

where j is a half-integral or integral non-negative number and $m=-j,-j+1,\ldots,j-1,j$. We have

$$\langle jm|\hat{\rho}^{(j)}|jm'\rangle = \rho^{(j)}_{mm'},$$

$$\hat{\rho}^{(j)} = \sum_{m=-j}^{j} \sum_{m'=-j}^{j} \rho^{(j)}_{mm'} |jm\rangle\langle jm'|. \tag{2}$$

The diagonal elements of the density matrix specify the distribution function of the spin along the z axis (in states with a density operator $\hat{\rho}^{(j)}$),

$$\rho^{(j)}_{mm} = \omega_0(m), \tag{3}$$

with the normalization condition

$$\sum_{m=-j}^{j} \omega_0(m) = 1. \tag{4}$$

In a reference frame rotated relative to the initial frame such that the rotation is specified by the Euler angles α, β, and γ, the distribution of the projections m_1 of the spin on the new axis z' is given by

$$\tilde{\omega}(m_1,\alpha,\beta,\gamma) = \sum_{m_1'=-j}^{j} \sum_{m_2'=-j}^{j} D^{(j)}_{m_1 m_1'}(\alpha,\beta,\gamma) \times \rho^{(j)}_{m_1' m_2'} D^{(j)*}_{m_1 m_2'}(\alpha,\beta,\gamma). \tag{5}$$

In this formula the matrix elements of the irreducible representation of the rotation group are specified by the Wigner $D^{(j)}$ function,[3]

$$D^{(j)}_{m'm}(\alpha,\beta,\gamma) = e^{im'\gamma} d^{(j)}_{m'm}(\beta) e^{im\alpha}, \tag{6}$$

where

$$d^{(j)}_{m'm}(\beta) = \left[\frac{(j+m')!(j-m)!}{(j+m)!(j-m)!}\right]^{1/2} \times \left(\cos\frac{\beta}{2}\right)^{m'+m} \left(\sin\frac{\beta}{2}\right)^{m'-m} \times P^{(m'-m,m'+m)}_{j-m'}(\cos\beta) \tag{7}$$

and $P^{(a,b)}_n(x)$ is the Jacobi polynomial.

The Wigner function $D^{(j)}_{m',m}$ has the property

$$D^{(j)*}_{m'm}(\alpha,\beta,\gamma) = (-1)^{m'-m} D^{(j)}_{-m'-m}(\alpha,\beta,\gamma). \tag{8}$$

Since in Eq. (5) we have the product of a matrix element of the representation of the rotation group and the complex conjugate matrix element belonging to the same row, it follows from Eq. (6) that there is no dependence on γ. Thus, the distribution function in Eq. (5) depends only on the two continuous rotation parameters α and β, i.e., it is a function of a point on the unit sphere. Hence, we introduce the notation

$$\omega(m_1,\alpha,\beta)=\tilde{\omega}(m_1,\alpha,\beta,\gamma) \tag{9}$$

for this distribution function and can write down the normalization conditions

$$\sum_{m_1=-j}^{j}\omega(m_1,\alpha,\beta)=1. \tag{10}$$

As an example, for $j=1/2$ and a state with spin along the z axis equal to $m_1=1/2$, the spin wave function is the spinor

$$\psi_{+1/2}=\begin{pmatrix}1\\0\end{pmatrix}$$

and the density matrix of this state has the form

$$\rho_{+}=\begin{pmatrix}1&0\\0&0\end{pmatrix}.$$

The distribution function associated with this state for $m_1=1/2$ is

$$\omega\left(\frac{1}{2},\alpha,\beta\right)=\cos^2\frac{\beta}{2}.$$

The corresponding value of this function for the state with a projected spin of $m_1=-1/2$ is

$$\omega\left(-\frac{1}{2},\alpha,\beta\right)=\sin^2\frac{\beta}{2}.$$

The distribution function (5) is a diagonal matrix element of the rotated spin density matrix (or density matrix of the spin state viewed from a rotated reference frame). Our problem is to obtain an expression for the density matrix of the state with matrix element $\rho^{(j)}_{mm'}$ in terms of the distribution function (9). To do this we multiply both sides of Eq. (5) by $D^{(j_3)}_{0m_3}(\alpha,\beta,\gamma)$ (i.e., by a spherical harmonic[3]) and integrate the equation over the Euler angles after first substituting for $D^{(j)*}_{m_1m_2'}(\alpha,\beta,\gamma)$ using Eq. (8). The spin j_3 for the function $D^{(j_3)}_{0m_3}(\alpha,\beta,\gamma)$ is an integer. Here we use the known integral of the product of three Wigner D-functions over the rotation group,[3]

$$\int D^{(j_1)}_{m_1'm_1}(\omega)D^{(j_2)}_{m_2'm_2}(\omega)D^{(j_3)}_{m_3'm_3}(\omega)\frac{d\omega}{8\pi^2}=\begin{pmatrix}j_1&j_2&j_3\\m_1'&m_2'&m_3'\end{pmatrix}\begin{pmatrix}j_1&j_2&j_3\\m_1&m_2&m_3\end{pmatrix}. \tag{11}$$

In this formula ω denotes the set of three Euler angles and

$$\int d\omega=\int_0^{2\pi}d\alpha\int_0^{\pi}\sin\beta\,d\beta\int_0^{2\pi}d\gamma. \tag{12}$$

The quantities on the right-hand side of Eq. (11) are the Wigner $3j$-symbols given explicitly in Ref. 3. We thus obtain

$$\int(-1)^{m_1}\omega(m_1,\alpha,\beta)D^{(j_3)}_{0m_3}(\alpha,\beta,\gamma)\frac{d\omega}{8\pi^2}=(-1)^{m_2'}\rho^{(j)}_{m_1'm_2'}\begin{pmatrix}j&j&j_3\\m_1&-m_1&0\end{pmatrix}\times\begin{pmatrix}j&j&j_3\\m_1'&-m_2'&m_3\end{pmatrix}. \tag{13}$$

Equation (13) is a linear system of equations for the unknown matrix elements $\rho^{(j)}_{m_1'm_2'}$. To solve this system we use the following properties of the $3j$-symbols:[3]

$$(2j+1)\sum_{m_1=-j_1}^{j_1}\sum_{m_2=-j_2}^{j_2}\begin{pmatrix}j_1&j_2&j\\m_1&m_2&-m\end{pmatrix}\times\begin{pmatrix}j_1&j_2&j'\\m_1&m_2&-m'\end{pmatrix}=\delta_{jj'}\delta_{mm'}, \tag{14}$$

and

$$\sum_{j=|j_1-j_2|}^{j_1+j_2}\sum_{m=-j}^{j}(2j+1)\begin{pmatrix}j_1&j_2&j\\m_1&m_2&-m\end{pmatrix}\times\begin{pmatrix}j_1&j_2&j\\m_1'&m_2'&-m\end{pmatrix}=\delta_{m_1m_1'}\delta_{m_2m_2'}. \tag{15}$$

Multiplying Eq. (13) by the $3j$-symbol

$$\begin{pmatrix}j&j&j_3\\m_1&-m_1&0\end{pmatrix}$$

and summing with the aid of Eq. (14) over the numbers m_1, we obtain

$$(2j_3+1)\sum_{m_1=-j}^{j}\int(-1)^{m_1}\omega(m_1,\alpha,\beta)D^{(j_3)}_{0m_3}(\alpha,\beta,\gamma)\times\begin{pmatrix}j&j&j_3\\m_1&-m_1&0\end{pmatrix}\frac{d\omega}{8\pi^2}=(-1)^{m_2'}\rho^{(j)}_{m_1'm_2'}\times\begin{pmatrix}j&j&j_3\\m_1'&-m_2'&m_3\end{pmatrix}. \tag{16}$$

Now multiplying Eq. (16) by the number

$$(2j_3+1)\begin{pmatrix}j&j&j_3\\m_1'&-m_2'&m_3\end{pmatrix}$$

and summing over the projected spins m_3 and the total spin j_3 using Eq. (15), we obtain

$$\sum_{j_3=0}^{2j}\sum_{m_3=-j_3}^{j_3}(2j_3+1)^2\times\sum_{m_1=-j}^{j}\int(-1)^{m_1}\omega(m_1,\alpha,\beta)D^{(j_3)}_{0m_3}(\alpha,\beta,\gamma)$$

$$\times\begin{pmatrix} j & j & j_3 \\ m_1 & -m_1 & 0 \end{pmatrix}\begin{pmatrix} j & j & j_3 \\ m_1' & -m_2' & m_3 \end{pmatrix}\frac{d\omega}{8\pi^2}$$
$$=(-1)^{m_2'}\rho^{(j)}_{m_1'm_2'}. \tag{17}$$

This relation is the solution of the stated problem. Since the function $D^{j_3}_{0m_3}(\alpha,\beta,\gamma)$ is a spherical harmonic that depends on only two of the Euler angles, α and β, Eq. (17) can be used to obtain an expression for the density matrix of the spin state in terms of an integral over the sphere of the product of the probability of the spin's having a projection m_1 in the direction normal to the sphere at the point with angular coordinates α,β and a spherical harmonic, followed by summation over the projections of the spin and the possible values of the total spin. Equation (17) has been obtained in other notation in Ref. 34. We have therefore shown that not only the density matrix, but also a marginal distribution can specify a mixed spin state. Consequently, we claim that a quantum mechanical spin state is specified, given the probability of the projection of the spin in a selected direction, $\omega(m_1,\alpha,\beta)$, measured in all arbitrarily rotated coordinate systems.

3. INVARIANT FORM OF SPIN TOMOGRAPHY

Equation (17) can be given an invariant operator form by using Eq. (2). We systematically introduce the following notation, first for the function on the unit sphere:

$$\Phi^{(j_3)}_{jm_1'm_2'}(\alpha,\beta)=(-1)^{m_2'}\sum_{m_3=-j_3}^{j_3} D^{(j_3)}_{0m_3}(\alpha,\beta,\gamma)\times\begin{pmatrix} j & j & j_3 \\ m_1' & -m_2' & m_3 \end{pmatrix}, \tag{18}$$

and then for the operator on the unit sphere:

$$\hat{A}^{(j_3)}_j(\alpha,\beta)=(2j_3+1)^2\sum_{m_1'=-j}^{j}\sum_{m_2'=-j}^{j}|jm_1'\rangle\times\Phi^{(j_3)}_{jm_1'm_2'}(\alpha,\beta)\langle jm_2'|. \tag{19}$$

In order to write a final expression for the density operator, we introduce an operator on the unit sphere which contains a dependence on the measurable projection of the spin:

$$\hat{B}^{(j)}_{m_1}(\alpha,\beta)=(-1)^{m_1}\sum_{j_3=0}^{2j}\begin{pmatrix} j & j & j_3 \\ m_1 & -m_1 & 0 \end{pmatrix}\hat{A}^{(j_3)}_j(\alpha,\beta). \tag{20}$$

Finally, we obtain an expression for the density operator,

$$\hat{\rho}^{(j)}=\sum_{m_1=-j}^{j}\int\frac{dw}{8\pi^2}\omega(m_1,\alpha,\beta)\hat{B}^{(j)}_{m_1}(\alpha,\beta). \tag{21}$$

This formula admits of the following interpretation. To determine the spin state for a spin j, one measures experimentally the projection m_1 of the spin in a direction specified by angles α and β, and for each direction the measurement yields a distribution function $\omega(m_1,\alpha,\beta)$ for a discrete set of projections of the spin $m_1=-j,-j+1,\ldots,j-1,j$. If this distribution function is known as a function of the parameters specifying a point on a sphere, then it is first used to find the averaged operator $\langle\hat{B}^{(j)}_{m_1}(\alpha,\beta)\rangle$ at each point of the unit sphere ($\langle\hat{B}^{(j)}_{m_1}(\alpha,\beta)\rangle$ is a basis operator that is independent of the measurement, although the density operator can be expressed in terms of it). Then the integral of the average operator over the solid angle is taken. Thus, we have

$$\hat{\rho}^{(j)}=\int\frac{dw}{8\pi^2}\langle\hat{B}^{(j)}_{m_1}(\alpha,\beta)\rangle, \tag{22}$$

where

$$\langle\hat{B}^{(j)}_{m_1}(\alpha,\beta)\rangle=\sum_{m_1=-j}^{j}\omega(m_1,\alpha,\beta)\hat{B}^{(j)}_{m_1}(\alpha,\beta). \tag{23}$$

Equation (22) completes the reconstruction of the density operator for the spin state. Using the operator $\langle\hat{B}^{(j)}_{m_1}\times(\alpha,\beta)\rangle$ averaged over the projection of the spin and specified at each point on the unit sphere, i.e., as a function of the angles α and β, we can write down the newly found density operator for the spin state in the following form:

$$\hat{\rho}^{(j)}=\frac{1}{4\pi}\int_0^{2\pi}d\alpha\int_0^{\pi}\sin\beta\,d\beta\langle\hat{B}^{(j)}_{m_1}(\alpha,\beta)\rangle. \tag{24}$$

These expressions make it possible to calculate the average values of physical quantities using the positive normalized distribution function $\omega(m_1,\alpha,\beta)$ to describe the quantum mechanical spin state. Thus, for any observable $\hat{K}$ the average value in a given spin state $\hat{\rho}^{(j)}$ can be found in the following way: let us apply the operator $\hat{K}$ to a function which does not depend on the quantum mechanical state:

$$\phi^{(j)}_K(m_1,\alpha,\beta)=\mathrm{Tr}[\hat{K}\hat{B}^{(j)}_{m_1}(\alpha,\beta)]. \tag{25}$$

Then the average value of the observable $\hat{K}$ in the state with distribution $\omega(m_1,\alpha,\beta)$ can be calculated according to the rule

$$\langle\hat{K}\rangle=\sum_{m_1=-j}^{j}\int\frac{dw}{8\pi^2}\omega(m_1,\alpha,\beta)\phi^{(j)}_K(m_1,\alpha,\beta). \tag{26}$$

Thus, each operator $\hat{K}$ corresponds to its symbol (25), which is analogous to the Weyl symbol and is a function of the discrete variable m_1 and the two continuous variables α and β. Subsequently, in specifying the state, the average value can be calculated as an average over the known distribution function followed by integration over the rotation parameters of the reference frame.

4. CONCLUSION

We have shown that, as for continuous variables like position, the quantum state of a spin degree of freedom can be specified by a positive probability distribution function that carries the same information on the spin state as the density matrix (density operator).

The principal results of this paper are Eqs. (21) and (24) for the density operator for a state with arbitrary spin j. These expressions serve as the basis of a system for specify-

ing and possibly measuring spin states using a positive normalized distribution function, which depends additionally on two angles that specify the direction along which the measured spin is projected. The result of the measurement of the state can be expressed as an angular integral of an expression that includes the measured distribution function, in complete analogy with an optical tomography scheme. Thus, in this paper we have discussed tomographic methods for measuring a quantum state described by a density operator for continuous variables and extended them to a purely quantum mechanical discrete variable, spin.

This work was supported by Grant No. 96-02-17222 of the Russian Fund for Fundamental Research.

[1] L. D. Landau, Z. Physik **45**, 430 (1927).
[2] E. Schrödinger, Ann. Phys. (Leipzig) **79**, 489 (1926).
[3] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Pergamon, Oxford (1977).
[4] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin (1932).
[5] L. de Broglie, Compt. Rend. **183**, 447 (1926); **184**, 273 (1927); **185**, 380 (1927).
[6] E. Madelung, Z. Phys. **40**, 332 (1926).
[7] D. Bohm, Phys. Rev. **85**, 166,180 (1952).
[8] E. Wigner, Phys. Rev. **40**, 749 (1932).
[9] R. J. Glauber, Phys. Rev. Lett. **10**, 84 (1963).
[10] E. C. G. Sudarshan, Phys. Rev. Lett. **10**, 277 (1963).
[11] K. Husimi, Proc. Phys. Math. Soc. Jpn. **23**, 264 (1940).
[12] Y. Kano, J. Math. Phys. **6**, 1913 (1965).
[13] J. Bertrand and P. Bertrand, Found. Phys. **17**, 397 (1987).
[14] K. Vogel and H. Risken, Phys. Rev. A **40**, 2847 (1989).
[15] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Phys. Rev. Lett. **70**, 1244 (1993).
[16] I. J. Dunn, I. A. Walmsley, and C. Mukamel, Phys. Rev. Lett. **74**, 884 (1995).
[17] S. Mancini, V. I. Man'ko, and P. Tombesi, Quantum Semiclassic. Opt. **7**, 615 (1995).
[18] G. M. D'Ariano, S. Mancini, V. I. Man'ko, and P. Tombesi, Quantum Semiclassic. Opt. **8**, 1017 (1996).
[19] G. M. D'Ariano, U. Leonhardt, and H. Paul, Phys. Rev. A **52**, R1801 (1995).
[20] S. Mancini, V. I. Man'ko, and P. Tombesi, Phys. Lett. A **213**, 1 (1996); E-prints archive, quant-ph/9609026; submitted to Found. Phys.
[21] S. Wallentowitz and W. Vogel, Phys. Rev. A **53**, 4528 (1996).
[22] K. Banaszek and K. Wodkiewicz, Phys. Rev. Lett. **76**, 4344 (1996).
[23] S. Mancini, V. I. Man'ko, and P. Tombesi, Europhys. Lett. **37**, 79 (1997).
[24] V. V. Dodonov, I. A. Malkin, and V. I. Man'ko, Physica **72**, 597 (1974).
[25] R. L. de Matos Filho and W. Vogel, Phys. Rev. A **54**, 4560 (1996).
[26] P. J. Bardroff, C. Leichte, G. Schrade, and W. P. Schleich, Phys. Rev. Lett. **77**, 2198 (1996).
[27] R. L. de Matos Filho and W. Vogel, Phys. Rev. Lett. **76**, 608 (1996).
[28] O. V. Man'ko, Preprint IC/96/39, ICTP, Trieste (1996); J. Russ. Laser Research **17**, 439 (1996).
[29] M. M. Nieto, Phys. Lett. A **219**, 180 (1996).
[30] O. V. Man'ko, Phys. Lett. A **228**, 29 (1997).
[31] D. M. Meekhof, G. Monroe, B. E. King, W. M. Itano, and D. Wineland, Phys. Rev. Lett. **76**, 1796 (1996).
[32] S. Haroche, Nuovo Cimento B **110**, 545 (1995).
[33] U. Leonhardt, Phys. Rev. A **53**, 2998 (1996).
[34] V. V. Dodonov and V. I. Man'ko, submitted to Phys. Lett. A.
[35] S. Mancini, V. I. Man'ko, and P. Tombesi, submitted to J. Mod. Opt.
[36] V. I. Man'ko, J. Russ. Laser Research **17**, 579 (1996).
[37] V. I. Man'ko, in *Symmetries in Science IX*, B. Gruber and M. Ramek, (eds.), Plenum, New York (1997), p. 215.

Translated by D. H. McNeill

Radiative corrections for level widths in light muonic atoms

V. G. Ivanov

Main Astronomical Observatory, Russian Academy of Sciences, 196140 St. Petersburg, Russia

S. G. Karshenboĭm

D. I. Mendeleyev Scientific-Research Institute of Metrology, 198005 St. Petersburg, Russia
(Submitted 25 February 1997)
Zh. Éksp. Teor. Fiz. **112**, 805–817 (September 1997)

The corrections, which are associated with electron vacuum polarization, for the radiation level widths and line intensities in light muonic atoms are examined. The total level widths, with allowance for the finite size of the nuclei, relativistic effects, and recoil are found. © *1997 American Institute of Physics.* [S1063-7761(97)00309-0]

1. INTRODUCTION

The radiative corrections of relative order $\alpha(Z\alpha)^2$ to the dipole matrix elements and the radiation level widths were examined in Refs. 1–3 in the logarithmic approximation. As is known, the leading radiative correction in muonic atoms is related to electron polarization and, being of relative order α, leads to corrections for the radiation level widths and line intensities amounting to several percentage points. As Z grows, the distribution of electric charge in the nucleus begins to play a noticeable role.

The difference between the leading radiative corrections in muonic atoms and those in ordinary atoms is due to the different characteristic atomic momenta. In ordinary light atoms, the atomic momenta $Z\alpha m_e$ are much smaller than the electron mass m_e, with the result that the contribution of the electron vacuum polarization is of relative order $\alpha(Z\alpha)^2$. The contribution of the electron's self-energy is of the same order but contains a large logarithm $\log(Z\alpha)^2$, so that in calculating the logarithmic approximation we can ignore the contribution of vacuum polarization. In muonic atoms the characteristic atomic momentum $Z\alpha m_\mu$ proves to be of the same order of magnitude as the electron mass m_e, so that electron polarization is predominant.

In this paper we discuss the corrections for the level widths of light muonic atoms with a nuclear charge $Z \leqslant 10$. The calculations are done for hydrogen-like atoms, but in the case of ions or neutral atoms, including electrons, the corrections related to the electron–muon interaction are small and, if necessary, can be taken into account separately.

Recall that, in contrast to ordinary atoms, in muonic systems the fine structure is smaller than the Lamb shifts, which are of a nonrelativistic nature if Z is low, so that the results for the radiative corrections can be found in the nonrelativistic approximation, with the relativistic corrections calculated separately.

The paper is organized as follows. First, we discuss the various estimates of radiative corrections, and then calculate them explicitly. The numerical results are listed for the $2p$ levels. The bulk of the paper deals with radiative corrections, but in the concluding part of the paper we discuss the total level widths and review all known and unknown contributions.

2. PRELIMINARY ESTIMATES FOR THE RADIATIVE CORRECTIONS

2.1. The delta-function-potential approximation

We start by reformulating the results of our earlier papers[1–3] in terms that are more convenient for investigating what follows. In these papers we found the radiative corrections of relative order $\alpha(Z\alpha)^2 \log(Z\alpha)$ to the probability of transitions between levels in hydrogen-like systems. The logarithmic contribution arises when we allow for the self-energy of the electron in the Coulomb field. The properties of this energy, when it acts as perturbation, differ considerably from the properties of vacuum polarization, which determines the leading radiative correction in the muonic atom. However, reasoning from the computational angle, we note that the results of Refs. 1–3 are based on the fact that the coefficient of the logarithm is a delta function and that there are no contributions related to the radiative corrections for the radiation and for the wave function of the p state. Another fact that plays an important role in simplifying the calculations is the absence of energy-dependence in the perturbation. If all these conditions are met, we need only allow for the corrections for the energy and the wave function of the s state, which are attributable to the single-loop operator of the self-energy of an electron (muon) in the Coulomb field of the nucleus.

A perturbation with a delta-function potential,

$$V(\mathbf{r}) = \delta(\mathbf{r}) \frac{\Delta E(1s)}{(\psi_{1s}(0))^2}, \tag{1}$$

leads to corrections for the energy,

$$\Delta E(nl) = \frac{\delta_{l0}}{n^3} \Delta E(1s), \tag{2}$$

and for the wave function of the nlm state,

$$\Delta\psi_{nlm}(\mathbf{r}) = \delta_{l0} \frac{R_{n0}(0)}{R_{10}(0)} \sum_{q \neq n} \psi_{q00}(\mathbf{r}) \frac{R_{q0}(0)}{R_{10}(0)} \frac{\Delta E(1s)}{E_{ns} - E_{qs}}, \tag{3}$$

where $R_{nl}(r)$ is the radial part of the Schrödinger wave function,

$$\psi_{nlm}(\mathbf{r}) = R_{nl}(r) Y_{lm}(\vartheta, \varphi), \tag{4}$$

 1063-7761/97/090435-06$10.00

and the sum incorporates all the intermediate states of the discrete and continuous spectra.

Allowing for these corrections, we find[1–3] the correction for the width of the $2p$ level

$$\Delta\Gamma_{2p}^{DP} = -(0.98\ldots)\mathscr{R}\Gamma^{(0)}(2p), \tag{5}$$

where the nonrelativistic width in the dipole approximation is given by the expression

$$\Gamma^{(0)}(i\to f) = \frac{4\omega_{fi}^3}{3}|\mathbf{d}_{fi}|^2 \tag{6}$$

and for a $2p\to 1s$ transition it is

$$\Gamma^{(0)}(2p) = \frac{2^9}{3^8}\alpha(Z\alpha)^2 E_0. \tag{7}$$

The effective parameter is defined as the shift of the $1s$ level,

$$\mathscr{R} = \frac{\Delta E_{1s}}{E_0}, \tag{8}$$

in units of the unperturbed energy of the ground state of the muonic atom (an analog of the Rydberg constant)

$$E_0 = \frac{(Z\alpha)^2 m_R}{2},$$

and m_R is the reduced mass. In all expressions we use the relativistic system of units $\hbar = c = 1$ and $\alpha = e^2$, and the phases of the wave functions of the s states in the coordinate representation are defined in such a way that their values at zero are real and positive.

In the first approximation the perturbative series in low-Z hydrogen-like atoms corresponds, as we know, to a delta-function potential. For instance, the distribution of the nuclear charge can be taken into account by the perturbation

$$V_{NC}(\mathbf{r}) = \frac{2\pi(Z\alpha)}{3}\langle R^2\rangle\,\delta(\mathbf{r}),$$

where $\langle R^2\rangle$ is the mean square of the charge radius of the nucleus. Thus, the result (5) can be used to allow for the charge distribution in the nucleus with a coefficient

$$\Delta E_{NC}(1s) = \frac{2}{3}(Z\alpha)^4 m_R^3\langle R^2\rangle. \tag{9}$$

The interaction responsible for nuclear polarization in light muonic atoms can also be assumed to be a delta function and allowed for by the correction (5).

In what follows we give a general expression for the radiative corrections for the level width and calculate these corrections explicitly, but before doing this we give an estimate of the contribution via (5). Indeed, for light muonic atoms, in the leading order in the parameter $Z\alpha$ there is still no energy-dependence of the perturbation and there are no corrections for the photon creation operator, but there are additional corrections for the wave function and the energy of the p states. However, it is quite obvious that the contributions to the shifts of the $1s$, $2s$, and $2p$ levels in light muonic atoms exhibit the following property (see, e.g., the results for muonic hydrogen[4,5] and muonic helium[6]): the results for the $2p$ levels are much smaller than for the $2s$ level, and the shifts of the $1s$ and $2s$ levels differ by a factor of 8 (n^3). Such energy shifts are characteristic of delta-function perturbations. On the other hand, the contribution to the correction for the dipole matrix element for a delta-function perturbation is largely determined by the off-diagonal matrix elements of the perturbation taken between the states $1s$ and $2s$ (see Refs. 1–3). In summary, we can estimate the correction for the width of the $2p$ levels using for the coefficient in (5) the numerical values of the Lamb shift of the ground level. The results for the vacuum polarization contribution can be conveniently written in the form

$$\Delta\Gamma = C_{VP}\mathscr{R}_{VP}\Gamma^{(0)}. \tag{10}$$

The effective parameters for various values of Z and the estimates are listed in Table I. The error in the estimates is characterized by the quantities

$$\kappa_1 = \frac{\Delta E(1s) - 8\Delta E(2s)}{\Delta E(1s)}, \quad \kappa_2 = \frac{\Delta E(2p)}{\Delta E(2s)},$$

which vanish in the case of a delta-function potential.

When estimating the errors, we must bear in mind that the final result (5) emerges[1–3] after large numerical reductions. The separate contributions related to the corrections for the frequency and the dipole matrix element in (6) are listed in Table II.

TABLE I. Estimates of single-loop radiative corrections for the width of the $2p\to 1s$ decay in the delta-function-potential approximation.

Z	$\mathscr{R}_{VP}/\alpha$	κ_1,%	κ_2,%	$\Delta\Gamma^{DP}/\alpha\Gamma^{(0)}$	$\Delta I^{DP}/\alpha I^{(0)}$
1	−0.10	8	7	0.10	0.24
2	−0.24	12	20	0.23	0.54
3	−0.34	11	31	0.33	0.78
4	−0.42	10	40	0.41	0.97
5	−0.49	8	46	0.48	1.1
6	−0.55	5	52	0.54	1.3
7	−0.60	3	56	0.59	1.4
8	−0.65	1	59	0.64	1.5
9	−0.69	−1	62	0.68	1.6
10	−0.73	−2	64	0.72	1.7

2.2. The effective charge approximation

Clearly, the estimates improve in accuracy with decreasing value of Z. However, as the nuclear charge grows, it becomes possible to estimate the contribution by employing what is known as the running coupling constant. Assuming that

TABLE II. Separate contributions found in the delta-function-potential (*DP*) approximation and the effective-charge approximation [the running coupling constant (*RC*)].

Quantity	Corrections for the frequency	Corrections for the matrix elements	Total contribution
$\Delta\Gamma^{DP}/\mathscr{R}_{VP}\Gamma^{(0)}$	−4.00	3.02	−0.98
$\Delta I^{DP}/\mathscr{R}_{VP}I^{(0)}$	−5.33	3.02	−2.32
$\Delta\Gamma^{RC}/\mathscr{R}_{VP}\Gamma^{(0)}$	−3.00	1.00	−2.00
$\Delta I^{RC}/\mathscr{R}_{VP}I^{(0)}$	−4.00	1.00	−3.00

TABLE III. Estimates of single-loop radiative corrections for the width of the $2p\to 1s$ decay in the running coupling constant approximation.

Z	$(1+\kappa_1)/2$, %	$1-\kappa_2$, %	$\Delta\Gamma^{RC}/\alpha\Gamma^{(0)}$	$\Delta I^{RC}/\alpha I^{(0)}$
3	55	69	0.67	1.0
4	55	60	0.84	1.2
5	54	54	1.0	1.5
6	52	48	1.1	1.7
7	51	44	1.2	1.8
8	50	41	1.3	2.0
9	50	38	1.4	2.1
10	49	36	1.4	2.2

$$Z\alpha m_\mu \gg m_e, \tag{11}$$

we find at a general rule for accounting for single-loop corrections:

$$Z\alpha(q)=Z\alpha\left\{1+\frac{\alpha}{\pi}\left(\frac{1}{3}\log\frac{q^2}{m_e^2}-\frac{5}{9}\right)\right\}, \tag{12}$$

where the effective momentum is defined as $q=Z\alpha m_\mu$.

Strictly speaking, allowing for a nonlogarithmic term in the expression for the running coupling constant gives better than expected accuracy and, additionally, for different states the effective momenta should differ slightly. Note that all nonrelativistic[1)] dipole radiation level widths are proportional to the same factor $(Z\alpha)^4$, so that there is no dependence on the states in this approximation. Note also that even at $Z=10$ the expressions for the corrections are not in the asymptotic region, which is evident from the expressions for the energy of the lowest levels (see below). Hence, instead of directly applying the rule (12), it is advisable to employ the fact that level widths and energies are related. To this end, we note that the level energies are proportional to $(Z\alpha)^2$, so that in the adopted approximation the quantities

$$\frac{\Delta E(1s)-4\Delta E(2s)}{\Delta E(1s)}=\frac{1+\kappa_1}{2},$$

$$\frac{\Delta E(2s)-\Delta E(2p)}{\Delta E(2s)}=1-\kappa_2$$

must vanish. We can write the correction as follows:

$$\Gamma_{2p}^{RC}=-2\mathscr{R}_{VP}\Gamma^{(0)}(2p), \tag{13}$$

and its accuracy is determined by the smallness of the quantities $(1+\kappa_1)/2$ and $1-\kappa_2$. Estimates for different values of Z are listed in Table III and the values of separate terms, in Table II.

To complete the discussion on estimates, we note that usually wave functions are more sensitive to perturbations than energies, so that the errors can be large. However, the fact that the results corresponding to opposite asymptotic conditions[2)] lead to quantities that differ only by a factor no greater than 2 suggest a weak dependence of the results on Z after the scale factor $\mathscr{R}$ has been isolated.

3. CALCULATION OF THE RADIATIVE CORRECTIONS

The radiative corrections for the energy levels and wave functions and, hence, for the level widths can be described nonrelativistically by employing the Uehling potential (provided that Z is moderately large). Here the width, with allowance for the correction, is described by Eq. (6) with the wave functions and energies found with allowance for the perturbing potential. To calculate the corrections for the level widths and intensities we must find the corrections for the dipole matrix elements and transition energies.

The expression for the leading radiative corrections of the order of α for the dipole matrix element in the muonic atom has the form

$$\Delta d_z(2p\to 1s)=\langle\Delta^{1s}|ez|2p\rangle+\langle 1s|ez|\Delta^{2p}\rangle,$$

where the corrections for the wave functions,

$$\langle\Delta^{nlm}|=\sum_{q\neq n}\frac{\langle nlm|V_{VP}|qlm\rangle}{E_n-E_q}\langle qlm|,$$

include the sum over all the intermediate states of the discrete and continuous spectra.

In terms of the normalized dipole matrix elements[2]

$$\mathscr{D}_q^s=\frac{d_z(2p\to qs)}{d_z(2p\to 1s)}\frac{R_{10}(0)}{R_{q0}(0)}, \tag{14}$$

$$\mathscr{D}_q^p=\frac{d_z(qp\to 1s)}{d_z(2p\to 1s)}\frac{R'_{21}(0)}{R'_{q1}(0)}, \tag{15}$$

the relative correction assumes the form

$$\frac{\Delta d_z(2p\to 1s)}{d_z(2p\to 1s)}=\sum_{q\neq 1}\mathscr{D}_q^s\frac{R_{q0}(0)}{R_{10}(0)}\frac{\langle 1s|V_{VP}|qs\rangle}{E_1-E_q}+\sum_{q\neq 2}\mathscr{D}_q^p\frac{R'_{q1}(0)}{R'_{21}(0)}\frac{\langle qp|V_{VP}|2p\rangle}{E_2-E_q}. \tag{16}$$

The explicit expressions for the dipole matrix elements needed for further calculations can be found in the Appendix.

The off-diagonal matrix element of electron vacuum polarization can easily be expressed in terms of the radial wave functions:

$$\langle nlm|V_{VP}(r)|ql'm'\rangle=-(Z\alpha)\frac{\alpha}{\pi}\delta_{mm'}\delta_{ll'}V_{nq}^l, \tag{17}$$

where

$$V_{nq}^l=\int_0^\infty dr\ r^2R_{nl}(r)R_{ql}(r)\int_0^1 dv\,\frac{v^2(1-v^2/3)}{1-v^2}\,\frac{e^{-\lambda r}}{r}, \tag{18}$$

$$\lambda=\frac{2m_e}{\sqrt{1-v^2}}. \tag{19}$$

Plugging in the explicit expressions for the wave functions (see the Appendix), we find

$$\int_0^\infty dr\ r\,R_{10}(r)R_{q0}(r)e^{-\lambda r}=\frac{R_{10}(0)R_{q0}(0)}{[\lambda+\gamma(1+1/q)]^2}\left[\frac{\lambda+\gamma(1-1/q)}{\lambda+\gamma(1+1/q)}\right]^{q-1}, \tag{20}$$

TABLE IV. Results of direct calculations of the width of the $2p$ level and the intensity of the $2p\to1s$ line. The upper index (s or p) indicates the level for which the correction for the wave function (Ψ) or the energy (E) is calculated.

Z	C_E^s	C_E^p	C_Ψ^s	C_Ψ^p	C_{VP}	$\Delta I_{VP}^{\rm tot}/\mathscr{R}_{VP}I^{(0)}$	$\Delta\Gamma_{VP}^{\rm tot}/\alpha\Gamma^{(0)}$	$\Delta I_{VP}^{\rm tot}/\alpha I^{(0)}$
1	−4.0	0.03	2.9	−0.09	−1.2	−2.5	0.12	0.26
2	−4.0	0.09	2.8	−0.22	−1.4	−2.7	0.32	0.63
3	−4.0	0.14	2.7	−0.31	−1.5	−2.8	0.51	0.94
4	−4.0	0.18	2.6	−0.37	−1.6	−2.9	0.66	1.2
5	−4.0	0.21	2.6	−0.42	−1.6	−2.9	0.80	1.4
6	−4.0	0.24	2.5	−0.46	−1.7	−2.9	0.93	1.6
7	−4.0	0.27	2.5	−0.49	−1.7	−3.0	1.0	1.8
8	−4.0	0.29	2.5	−0.51	−1.8	−3.0	1.1	1.9
9	−4.0	0.31	2.4	−0.53	−1.8	−3.0	1.2	2.1
10	−4.0	0.33	2.4	−0.54	−1.8	−3.0	1.3	2.2

$$\int_0^\infty dr\ r\ R_{21}(r)R_{q1}(r)e^{-\lambda r}$$

$$=\frac{96R'_{21}(0)R'_{q1}(0)}{[2\lambda+\gamma(1+2/q)]^4}\left[\frac{2\lambda+\gamma(1-2/q)}{2\lambda+\gamma(1+2/q)}\right]^{q-2}. \quad (21)$$

As a result, calculation of the matrix element (17) reduces to evaluating one-dimensional integrals:

$$V^s_{1q}=R_{10}(0)R_{q0}(0)\int_0^1 dv\ \frac{v^2(1-v^2/3)}{1-v^2}$$

$$\times\frac{1}{[\lambda+\gamma(1+1/q)]^2}\left[\frac{\lambda+\gamma(1-1/q)}{\lambda+\gamma(1+1/q)}\right]^{q-1}, \quad (22)$$

$$V^p_{q2}=R'_{21}(0)R'_{q1}(0)\int_0^1 dv\ \frac{v^2(1-v^2/3)}{1-v^2}$$

$$\times\frac{96}{[2\lambda+\gamma(1+2/q)]^2}\left[\frac{2\lambda+\gamma(1-2/q)}{2\lambda+\gamma(1+2/q)}\right]^{q-2}. \quad (23)$$

The final expression for the contribution of single-loop vacuum polarization to the width of the $2p$ level is

$$\frac{\Delta d_z(2p\to1s)}{d_z(2p\to1s)}$$

$$=\frac{\alpha}{\pi}\left\{8\sum_{n\neq1}\frac{1}{n^3}\frac{\mathscr{D}^s_n\widetilde{V}^s_{1n}}{1-1/n^2}\right.$$

$$+8\int_0^\infty\frac{dt}{t^3(1+1/t^2)}\frac{\mathscr{D}^s_t\widetilde{V}^s_{1t}}{1-e^{-2\pi t}}$$

$$+\frac{1}{12}\sum_{n\neq2}\frac{32}{n^5}\frac{n^2-1}{3}\frac{\mathscr{D}^p_n\widetilde{V}^p_{n2}}{1/4-1/n^2}$$

$$\left.+\frac{1}{12}\int_0^\infty\frac{32dt}{t^5(1/4+1/t^2)}\frac{1+t^2}{3}\frac{\mathscr{D}^p_t\widetilde{V}^p_{t2}}{1-e^{-2\pi t}}\right\}, \quad (24)$$

where the following notation is introduced:

$$V^s_{1q}=\frac{R_{10}(0)R_{q0}(0)\widetilde{V}^s_{1q}}{\gamma^2},\quad V^p_{q2}=\frac{R'_{21}(0)R'_{q1}(0)\widetilde{V}^p_{q2}}{\gamma^2},$$

which is more convenient for analytic continuation to the states of the continuous spectrum. The expressions for $\widetilde{V}^s_{1t}$ and $\widetilde{V}^p_{t2}$ can be obtained by replacing n with $-it$.

Numerical integration leads to

$$\frac{\Delta d_z(2p\to1s)}{d_z(2p\to1s)}=\frac{\alpha}{\pi}[-0.5867+0.0805+0.0525]$$

$$=-0.4537\,\frac{\alpha}{\pi}$$

for muonic hydrogen and

$$\frac{\Delta d_z(2p\to1s)}{d_z(2p\to1s)}=\frac{\alpha}{\pi}[-1.274+0.199+0.135]$$

$$=-0.940\,\frac{\alpha}{\pi}$$

for muonic helium.

The terms in the square brackets originate from the contributions of states with $n=2$, higher bound states, and the continuous spectrum, respectively. We see that numerically the contribution with $n=2$ is predominant, which justifies the estimates with formulas for a delta-function potential.

The results for $Z\leqslant10$ are given in Table IV. We see that the contributions from the p states are smaller then those from the s states, which justifies the estimates based on the use of a delta-function potential. We then see that both estimates given in Sec. 2 (see Tables I–III) agree with the results of direct calculations. The agreement between the estimates and the results of direct calculations suggests that similar estimates can be applied to level widths and intensities of the transitions involving higher levels. The results for a delta-function potential for transitions between levels with $n,n'\leqslant4$ needed for estimates have been obtain in Ref. 3.

TABLE V. The coefficients $C_{\rm Rel}$ and $C'_{\rm Rel}$ [see Eq. (25)] for calculating the relativistic correction $\Delta\Gamma_{\rm Rel}$.

State	$C_{\rm Rel}$	$C'_{\rm Rel}$
$2p_{1/2}$	$\ln\frac{9}{8}$	1
$2p_{3/2}$	$-\frac{7}{48}-\frac{1}{2}\ln\frac{32}{27}$	1

TABLE VI. The leading contributions to the width $\Gamma(2p_j)$ [see Eq. (25)]. The recoil corrections ($\Delta\Gamma_{\rm Rec}$) emerge because of the difference between the factor $(m_R/m_\mu)^2(1+Zm_\mu/M)^2$ and unity.

					$\Delta\Gamma_{\rm Rel}/\Gamma^{(0)}$	
z	A	$\Delta\Gamma_{\rm Rec}/\Gamma^{(0)}$	$\Delta\Gamma_{VP}/\Gamma^{(0)}$	$\Delta\Gamma_{NC}/\Gamma^{(0)}$	$j=1/2$	$j=3/2$
1	1	0	$8.729(9)\times10^{-4}$	$-1.20(3)\times10^{-5}$	1.11×10^{-5}	-7.4×10^{-6}
2	4	0.0558903	$2.349(4)\times10^{-3}$	$-2.12(3)\times10^{-4}$	3.65×10^{-5}	-3.77×10^{-5}
3	7	0.0646677	$3.682(5)\times10^{-3}$	$-1.00(3)\times10^{-3}$	7.90×10^{-5}	-8.81×10^{-5}
4	9	0.075988	$4.840(3)\times10^{-3}$	$-1.98(8)\times10^{-3}$	1.422×10^{-4}	-1.548×10^{-4}
5	11	0.083268	$5.858(3)\times10^{-3}$	$-2.8(2)\times10^{-3}$	2.240×10^{-4}	-2.401×10^{-4}
6	12	0.095858	$6.753(2)\times10^{-3}$	$-4.3(2)\times10^{-3}$	3.326×10^{-4}	-3.359×10^{-4}
7	14	0.098775	$7.554(3)\times10^{-3}$	$-6.3(4)\times10^{-3}$	4.531×10^{-4}	-4.569×10^{-4}
8	16	0.10104	$8.262(2)\times10^{-3}$	$-9.3(8)\times10^{-3}$	5.923×10^{-4}	-5.965×10^{-4}
9	19	0.09724	$8.913(3)\times10^{-3}$	$-0.013(1)$	7.374×10^{-4}	-7.674×10^{-4}
10	20	0.1042	$9.504(3)\times10^{-3}$	$-0.018(2)$	9.265×10^{-4}	-9.317×10^{-4}

4. CONCLUSIONS

Now let us discuss the expression for the total radiation width of the $2p_j$ levels:

$$\Gamma_{2p_j}(Z)=\Gamma^{(0)}(2p_j)\left(\frac{m_R}{m_\mu}\right)^2\left(1+Z\frac{m_\mu}{M}\right)^2\Bigg[1+C_{\rm Rel}(2p_j)\times(Z\alpha^2)+C'_{\rm Rel}(2p_j)\,\frac{Zm_\mu}{M}(Z\alpha)^2+C_{VP}(2p).\mathscr{R}_{VP}+C_{NC}(2p).\mathscr{R}_{NC}\Bigg]. \tag{25}$$

Here the nonrelativistic dipole width $\Gamma^{(0)}$ has been defined in (7), and its numerical value can be found by plugging in the value of the ground-state energy

$$E_0=2813.226(3)Z^2\frac{m_R}{m_\mu}\ \text{eV}.$$

The effective parameters $\mathscr{R}$ and the constants C for the vacuum polarization contribution and the effects associated with the finiteness of the size of the nucleus have been considered earlier. Recoil effects in the nonrelativistic approximation are taken into account by a special factor,[7]

$$\left(\frac{m_R}{m}\right)^2\left(1+Z\frac{m_\mu}{M}\right)^2,$$

where M is the mass of the nucleus. The relativistic coefficient $C_{\rm Rel}$ can easily be found by calculating the width via the Dirac wave functions (the result for the $2p_{1/2}$ state agrees with that of Sokolov and Yakovlev[8] but disagrees with that of Borovskiĭ *et al.*;[9] in the case of the $2p_{3/2}$ state our result agrees with that of Ref. 9). The relativistic part of the recoil can also be easily taken into account (see Table V).

Table VI lists the most important contributions to the level widths. The largest is the correction induced by the single-loop electron vacuum polarization. The error in allowing for the finite size of the nucleus is related to higher-order corrections in $Z\alpha$. Note that the corrections under discussion depend differently on the nuclear charge Z, and smaller corrections may become noticeable in heavier muonic atoms.

Some researchers have performed highly accurate studies of the intensity ratios for lines in muonic atoms (see Ref. 10 and the papers cited there). Various corrections for the intensities of $2p\to1s$ lines are listed in Table VII. The calculation errors amount to about one percentage point, and the corrections discussed in this paper may become significant.

The final results for the widths are listed in Table VIII, with the errors determined by an estimate of the uncalculated higher-order contributions and by the accuracy of numerical integration of the vacuum polarization contribution (see Table VI).

Part of this work was sponsored by the Russian Fund for

TABLE VII. The leading corrections for the intensities of the lines of $2p_j\to1s$ transitions.

					$\Delta I_{\rm Rel}/I^{(0)}$	
Z	A	$\Delta I_{\rm Rec}/I^{(0)}$	$\Delta I_{VP}/I^{(0)}$	$\Delta I_{\rm NC}/I^{(0)}$	$j=1/2$	$j=3/2$
1	1	0	$1.8665(9)\times10^{-3}$	$4.3(6)\times10^{-6}$	1.57×10^{-5}	-1.2×10^{-6}
2	4	0.0558903	$4.588(4)\times10^{-3}$	$7.5(5)\times10^{-5}$	5.48×10^{-5}	-1.28×10^{-5}
3	7	0.0646677	$6.848(5)\times10^{-3}$	$3.6(5)\times10^{-4}$	1.202×10^{-4}	-3.19×10^{-5}
4	9	0.075988	$8.737(3)\times10^{-3}$	$7(1)\times10^{-3}$	2.154×10^{-4}	-5.50×10^{-5}
5	11	0.083268	0.010359(3)	$1.0(3)\times10^{-3}$	3.38×10^{-4}	-8.4×10^{-5}
6	12	0.095858	0.011763(2)	$1.5(4)\times10^{-3}$	4.97×10^{-4}	-1.11×10^{-4}
7	14	0.098775	0.013007(3)	$2.2(7)\times10^{-3}$	6.77×10^{-4}	-1.51×10^{-4}
8	16	0.10104	0.014104(2)	$3(1)\times10^{-3}$	8.9×10^{-4}	-2.0×10^{-4}
9	19	0.09724	0.015105(3)	$5(2)\times10^{-3}$	1.11×10^{-3}	-2.6×10^{-4}
10	20	0.1042	0.016008(3)	$6(4)\times10^{-3}$	1.38×10^{-3}	-3.1×10^{-4}

TABLE VIII. Total widths of the $2p_j \to 1s$ transitions.

Z	A	$\Gamma_{1/2}$	$\Gamma_{3/2}$
1	1	0.0767426(7)	0.0767412(7)
2	4	1.40457(3)	1.40446(3)
3	7	7.2599(3)	7.2587(3)
4	9	23.276(2)	23.269(2)
5	11	57.36(1)	57.32(1)
6	12	120.36(3)	120.28(3)
7	14	223.6(1)	223.4(1)
8	16	381.9(3)	381.4(3)
9	19	608.2(8)	607.2(8)
10	20	929(2)	928(2)

Fundamental Research (Grant No. 95-02-03977).

APPENDIX: EXPLICIT EXPRESSIONS FOR WAVE FUNCTIONS AND MATRIX ELEMENTS

The radial parts of the wave functions (4) have the form

$$R_{10}(r)=R_{10}(0)\exp(-\gamma r),$$

$$R_{n0}(r)=R_{n0}(0)\exp\left(-\frac{\gamma r}{n}\right)F(1-n,2;2\gamma r/n),$$

$$R_{k0}(r)=R_{k0}(0)\exp(-ikr)F(1+i\gamma/k,2;2ikr),$$

$$R_{21}(r)=R'_{21}(0)\,\gamma r\,\exp\left(-\frac{\gamma r}{2}\right),$$

$$R_{n1}(r)=R'_{n1}(0)\,\frac{2\gamma r}{n}\exp\left(-\frac{\gamma r}{n}\right)F(2-n,4;2\gamma r/n),$$

$$R_{k1}(r)=R'_{k1}(0)2kr\,\exp(-ikr)F(2+i\gamma/k,4;2ikr),$$

where $\gamma=Z\alpha m_R$, $F(a,b;z)$ is the confluent hypergeometric function, and the wave number k corresponds to states of the continuous spectrum.

The radial parts of the wave functions are defined in such a way that their values at zero for s states and the values of the derivatives for p states,

$$R'_{q1}(0)=\frac{1}{\gamma}\left.\frac{\partial R_{q1}(r)}{\partial r}\right|_{r=0}, \tag{A1}$$

are real and positive, and in the case of bounds states are

$$\left(\frac{R_{n0}(0)}{R_{10}(0)}\right)^2=\frac{1}{n^3} \tag{A2}$$

and

$$\left(\frac{R'_{n1}(0)}{R'_{21}(0)}\right)^2=\frac{32}{3}\,\frac{1}{n^3}\left(1-\frac{1}{n^2}\right), \tag{A3}$$

respectively. For states belonging to the continuous spectrum, the above equalities become

$$\left(\frac{R_{k0}(0)}{R_{10}(0)}\right)^2=\frac{2\pi}{\gamma}\,\frac{k/\gamma}{1-\exp(-2\pi\gamma/k)} \tag{A4}$$

and

$$\left(\frac{R'_{k1}(0)}{R'_{21}(0)}\right)^2=\frac{32}{3}\,\frac{2\pi}{\gamma}\,\frac{k^3/\gamma}{1-\exp(-2\pi\gamma/k)}\left(1+\frac{\gamma^2}{k^2}\right), \tag{A5}$$

respectively.

The normalized dipole matrix elements have the form

$$\mathscr{D}^{\,s}_{n}=3^4\frac{n^6}{(n^2-4)^3}\left(\frac{n-2}{n+2}\right)^n, \tag{A6}$$

$$\mathscr{D}^{\,p}_{n}=\frac{3^5}{2^6}\,\frac{n^6}{(n^2-1)^3}\left(\frac{n-1}{n+1}\right)^n \tag{A7}$$

for discrete states and

$$\mathscr{D}^{\,s}_{t}=3^4\frac{t^6}{(t^2+4)^3}\,\exp\left(-2t\,\arctan\frac{2}{t}\right), \tag{A8}$$

$$\mathscr{D}^{\,p}_{t}=\frac{3^5}{2^6}\,\frac{t^6}{(t^2+1)^3}\,\exp\left(-2t\,\arctan\frac{1}{t}\right) \tag{A9}$$

for continuous states. Here $t=\gamma/k$ is the continuous analog of the principal quantum number n.

[1)] In addition to the asymptotic condition (11), the range of applicability of the nonrelativistic estimate being discussed is restricted by the obvious condition $Z\alpha\ll 1$.

[2)] Clearly, a delta-function potential emerges in the $Z\alpha m_\mu \ll m_e$ limit.

[1] S. G. Karshenboĭm, Zh. Éksp. Teor. Fiz. **106**, 414 (1994) [JETP **79**, 230 (1994)]; Yad. Fiz. **58**, 901 (1995) [Phys. At. Nucl. **58**, 835 (1995)].
[2] S. G. Karshenboĭm, Zh. Éksp. Teor. Fiz. **107**, 1061 (1995) [JETP **80**, 593 (1995)].
[3] V. G. Ivanov and S. G. Karshenboĭm, Phys. Lett. A **210**, 313 (1996); Zh. Éksp. Teor. Fiz. **109**, 1219 (1996) [JETP **82**, 656 (1996)].
[4] A. D. Galanin and I. Ya. Pomeranchuk, Dokl. Akad. Nauk. SSSR **96**, 251 (1952).
[5] A. I. Akhiezer and V. B. Berestetskiĭ, *Quantum Electrodynamics*, Wiley, New York (1974).
[6] G. A. Rinker, Phys. Rev. A **14**, 18 (1976); L. W. Fullerton and G. A. Rinker, Phys. Rev. A **13**, 1283 (1976).
[7] Z. Fried and A. D. Martin, Nuovo Cimento **29**, 574 (1963).
[8] Yu. L. Sokolov and V. P. Yakovlev, Zh. Éksp. Teor. Fiz. **83**, 15 (1982) [Sov. Phys. JETP **56**, 7 (1982)].
[9] A. V. Borovskiĭ, S. A. Zapryagaev, O. I. Zatsarinnyĭ, and N. A. Manakov, *Plasma of Multiply Charged Ions* [in Russian], Khimiya, St. Petersburg (1995).
[10] C. J. Orth, M. E. Schilaci, and J. D. Knight, Phys. Rev. A **25**, 876 (1982).

Translated by Eugene Yankovsky
Edited by S. J. Amoretty

A modified Jaynes–Cummings model for an atom interacting with a classical multifrequency field

M. Z. Smirnov

Laser Center, St. Petersburg State Institute of Precision Mechanics and Optics, 197101 St. Petersburg, Russia
(Submitted 28 February 1997)
Zh. Éksp. Teor. Fiz. **112**, 818–827 (September 1997)

An open quantum system, which consists of a ''dressed'' two-level atom, i.e., an atom interacting with a classical multifrequency field, and a single quantized mode of an electromagnetic field, is examined. It is shown that when the frequency of the quantized mode coincides with one of the transition frequencies between the quasienergy levels, two interaction mechanisms, which differ in the dynamics of the populations of the quasienergy states, can be realized.
 [S1063-7761(97)00409-5]

1. INTRODUCTION

At present, there exists a large number of theoretical papers devoted to the study of the Jaynes–Cummings model, a closed quantum system consisting of a two-level atom and a single quantized mode of an electromagnetic field.[1] An analytic solution of the Heisenberg equations for the atom and field operators has been obtained in the dipole-interaction and rotating-wave approximations.[2–4] This solution makes it possible to demonstrate some fundamental effects emerging in more complicated systems, e.g., pulsations of atomic level populations and the generation of squeezed quantum states of a field.[3,5,6]

The nonlinear dynamics of the Jaynes–Cummings model can be qualitatively characterized in the following way. If initially the atom was in the ground quantum state and the quantized mode contained n photons, subsequently the atom and field will periodically exchange energy. The populations of the atomic levels and the dipole moment of the atom undergo pulsations, whose frequency (the generalized Rabi frequency) depends on n. If the initial number of photons n is not defined exactly in the quantum mechanical sense (say, initially the quantized mode is excited to a coherent state $|\alpha\rangle$), the pulsations are represented by a linear combination of oscillations with different frequencies and amount to periodic collapses and revivals (provided that the phases of the component oscillations are matched).[3,5] Oscillating solutions have also been obtained for more complicated closed systems generalizing the Jaynes–Cummings model.[7–10] Such systems incorporate a three- or four-level atom which interacts with one or several quantized modes of an electromagnetic field. Some solutions that do not use the rotating-wave approximation have also been derived.[11]

Lately, there has been an upsurge of interest in the experimental applications of the Jaynes–Cummings model.[7] In particular, experiments with a single-atom maser[12,13] made it possible to observe pulsations of the population of an excited state of an atom when the atom interacts with an electromagnetic field in a high-Q superconducting cavity. These experiments served as a powerful stimulus in the studies of generalizations of the Jaynes–Cummings model aimed at describing more complicated quantum systems.[14–18] We believe that a promising generalization of this model should incorporate open systems. In this connection, we would like to mention the work of Dung and Shumovski,[19] who studied the effect of relaxation on the fluctuations of the phase of the quantized field.

In this paper we introduce a modified Jaynes–Cummings model, i.e., a ''dressed'' atom that interacts with a single quantized mode of an electromagnetic field. According to modern terminology, a ''dressed'' atom is interpreted as an atom that interacts with an external classical field.[20] Such an atom is an open quantum system and its energy is not conserved. At the same time, if the classical field has an equidistant frequency spectrum, then for a ''dressed'' atom we can introduce the concept of quasienergy which, in a certain sense, is conserved[20–23] (we study the corresponding formalism in Sec. 2). Thus, a ''dressed'' two-level atom is characterized by an infinite number of quasienergy levels, whose positions depend on the parameters of the classical field. When such an atom interacts with a probe field[24,25] or with an electromagnetic-field vacuum (resonance fluorescence[26,27]), resonance effects caused by quantum transitions between quasienergy levels and the crossing (anticrossing) of such levels manifest themselves.

The plan of the paper is as follows. In Sec. 2 we use our earlier work[23,28] to briefly describe the properties of the quasienergy states of a ''dressed'' atom. The treatment of this problem differs from that used by other researchers in that we employ the Heisenberg picture, which we find more convenient in solving the Jaynes–Cummings problem. In Sec. 3 we solve a new problem of the interaction of a ''dressed'' atom and a quantized mode of an electromagnetic field. In Sec. 4, the last section of this article, we review the main results.

2. THE QUASIENERGY STATES OF A TWO-LEVEL ATOM

Let us take a two-level atom that interacts with a periodically modulated classical field

$$E(t)=\tilde{E}g(\omega' t)\exp(i\Omega t)+\text{c.c.}, \tag{1}$$

where $\tilde{E}$ is the complex-valued amplitude of the field, $g(\omega t)$ is the 2π-periodic (complex-valued) modulation function, ω'

is the modulation frequency, and Ω is the optical carrier frequency. If we use the dipole-interaction and the rotating-wave approximations,[4] we can write the Hamiltonian of the atom in the external field as follows:

$$\hat{H}_a=\hbar\omega\hat{b}^\dagger\hat{b}-\hbar\omega'\sigma g(\omega' t)\exp[i(\Omega t+\psi)]\hat{b}+\text{H.c.}, \quad (2)$$

where ω is the transition frequency between the atomic levels, $\sigma\equiv|\mu_{01}\tilde{E}/\hbar\omega|$ is the Rabi frequency normalized to the modulation frequency, with μ_{01} the matrix element of the dipole moment of the transition, $\psi\equiv\arg[\mu_{01}\tilde{E}/\hbar\omega']$, $\hat{b}=|0\rangle\langle 1|$ is the transition operator between the lower ($|0\rangle$) and upper ($|1\rangle$) energy levels of the atom, and H.c. stands for an expression that is the Hermitian conjugate of the first two terms on the right side of Eq. (2). The Schrödinger equation with the Hamiltonian (2) has two orthonormalized solutions, known as quasienergy states:[20,25]

$$|\theta_1\rangle=\exp(i\lambda\tau)[\varphi_1(\tau)|0\rangle+\exp(-i\Omega t)\varphi_2(\tau)|1\rangle],$$

$$|\theta_2\rangle=\exp[i(\delta-\lambda)\tau][\varphi_2^*(\tau)|0\rangle+\exp(-i\Omega t)\varphi_1^*(\tau)|1\rangle], \quad (3)$$

where $\tau=\omega' t$ is the dimensionless time variable, and $\delta=(\Omega-\omega)/\omega'$ is the detuning of the field frequency from the transition frequency, normalized to the modulation frequency. The value of the real quantity λ is determined (modulo 2) from the condition that the system of equations [obtained by plugging (3) into the Schrödinger equation]

$$\frac{d}{d\tau}\varphi_1=-i\lambda\varphi_1+i\sigma g(\tau)\varphi_2,$$

$$\frac{d}{d\tau}\varphi_2=-i(\delta-\lambda)\varphi_1+i\sigma g^*(\tau)\varphi_1 \quad (4)$$

has bounded 2π-periodic solutions. The introduction of the normalization condition

$$1=\langle\theta_1|\theta_1\rangle=\langle\theta_2|\theta_2\rangle=|\varphi_1(\tau)|^2+|\varphi_2(\tau)|^2$$

makes it possible to uniquely determine (to within a phase factor) the 2π-periodic functions $\varphi_1(\tau)$ and $\varphi_2(\tau)$ from the system of equations (4). Note that these results, obtained in the rotating wave approximation, are valid if

$$\omega',\lambda\omega',\delta\omega',\sigma\omega'\ll\Omega. \quad (5)$$

Physically, the quantity λ introduced above determines the quasienergy of the atom. When there is a classical field, each energy level of the atom splits into two equidistant sequences of quasilevels. The energies of these quasilevels are defined as[23]

$$\hbar\omega'(-\lambda+2n), \quad \hbar\omega'(\lambda-\delta+2m+1)$$

near the lower energy level of the isolated atom, and as

$$\hbar\omega+\hbar\omega'[\delta-\lambda-(2m+1)], \quad \hbar\omega+\hbar\omega'(\lambda-2n)$$

near the upper level, with $m,n=0,\pm1,\pm2,\ldots$. Such level splitting (dynamic Stark splitting) in a periodically modulated field is shown schematically in Fig. 1. Each quasilevel sequence has a period in energy, $2\hbar\omega'$, and is represented in Fig. 1 by a single quasilevel: $0'$, $1'$, $0''$, or $1''$. The quasilevels (quasilevel sequences) $0'$ and $0''$ correspond to the quasienergy state $|\theta_0\rangle$ and the quasilevels $1'$ and $1''$, to the state $|\theta_1\rangle$. Figure 1 shows that there are three transition frequencies between the quasilevels, which are determined (modulo $2\hbar\omega'$) by the values ω_1, ω_2, and ω_3. Spectroscopically, the quasilevel structure manifests itself in the same way as the system of energy levels of an isolated atom. In particular, in the resonance fluorescence spectra of a ''dressed'' atom one can observe resonance peaks at frequencies of the transitions between the quasilevels.[26,27] However, if we consider the nonlinear atom–field interaction, differences emerge between quasilevels and energy levels. These differences are examined in Sec. 3.

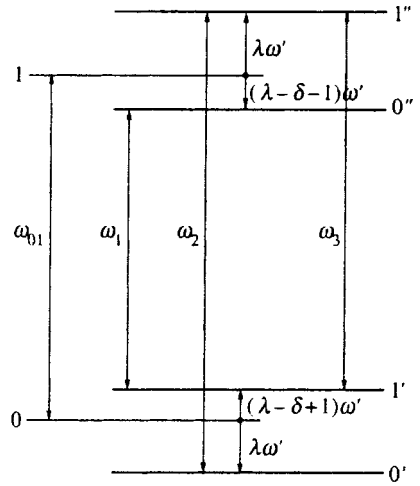

FIG. 1. The quasilevel structure of a ''dressed'' atom: 0 and 1 are the energy levels of an unperturbed atom, and $0'$, $1'$, $0''$, and $1''$ are the quasilevels (one quasilevel from each equidistant series is shown).

The properties of quasilevels and quasienergy states for various cases are discussed in Refs. 23–25 and 28. A generalization of the formalism, which we are discussing in this paper, to the case of a multilevel atom was proposed in Ref. 28. A different approach proposed by other researchers[20–22] has led to the same quasienergy values. However, the latter approach presupposes the introduction of an infinite set of ''Floquet states'' instead of two quasienergy states, which, we believe, complicates the analysis of the nonlinear interaction of a ''dressed'' atom and a quantized electromagnetic field.

Below we use the Heisenberg picture instead of the Schrödinger picture. The dynamics of a ''dressed'' atom is described by the Heisenberg equation

$$\frac{d}{dt}\hat{b}=-\frac{i}{\hbar}[\hat{b},\hat{H}_a]=-i\omega\hat{b}+i\omega'\sigma g^*(\omega' t)\times\exp[i(\omega t+\psi)](\hat{I}-2\hat{b}^\dagger\hat{b}), \quad (6)$$

where $\hat{I}$ is the identity operator. The transition operator between the quasienergy states can be written as

$$\hat{c}\equiv|\theta_0\rangle\langle\theta_1|=\exp[i(2\lambda-\delta)\tau][\varphi_1\varphi_2(\hat{I}-2\hat{b}^\dagger\hat{b})-\exp(i\Omega t')\varphi_1^2\hat{b}+\exp(-i\Omega t')\varphi_2^2\hat{b}^\dagger], \quad (7)$$

where $\Omega t\equiv\Omega t+\psi$, and we have used the quasienergy states (3). Taking the time derivative of (7) and combining the

result with (4) and (6), we find that $(d/dt)\hat{c}=0$, i.e., the operator $\hat{c}$ is a constant of motion of a ''dressed'' atom. Actually, this follows from the definition of quasienergy states as solutions of the Schrödinger equation with the Hamiltonian $\hat{H}_a$.

3. INTERACTION OF A ''DRESSED'' ATOM AND A MODE OF A QUANTIZED ELECTROMAGNETIC FIELD

An external disturbance (say, a thermostat or an electromagnetic field) acting on a ''dressed'' atom can, in general, induce quantum transitions between the quasienergy levels. Formally, this means that the operator $\hat{c}\equiv|\theta_0\rangle\langle\theta_1|$ ceases to be a constant of motion and its rate of variation in time becomes proportional to the intensity of the disturbance. In the Jaynes–Cummings problem, a quantized electromagnetic field acts as the external disturbance. We consider only one mode of the quantized field and write the field strength operator in the form

$$\hat{E}_q=i\mathbf{e}[u(\mathbf{r})\hat{a}(t)-u^*(\mathbf{r})\hat{a}^\dagger(t)], \tag{8}$$

where $\mathbf{e}$ is the unit polarization vector, $u(\mathbf{r})$ is the normalized eigenfunction, and $\hat{a}$ and $\hat{a}^\dagger$ are the photon annihilation and creation operators. The Hamiltonian of the interaction of the ''dressed'' atom and the quantized field (8) can be written as

$$\hat{V}_{aq}=i\hbar\omega'\kappa\hat{b}\hat{a}^\dagger+\text{H.c.}, \tag{9}$$

where we have used the ordinary dipole-interaction and rotating-wave approximations and introduced the coupling constant $\kappa=(\boldsymbol{\mu}_{01}\cdot\mathbf{e})u(\mathbf{r})/\hbar\omega'$. In the latter expression the vector $\mathbf{r}$ specifies the position of the atom. The total Hamiltonian of the modified Jaynes–Cummings model considered below is

$$\hat{H}=\hat{H}_a+\hbar\omega_q\hat{a}^\dagger\hat{a}+\hat{V}_{aq}, \tag{10}$$

where ω_q is the frequency of the quantized field. The Heisenberg equations for the atom and the quantized field can be written in the following manner:

$$\begin{aligned}\frac{d}{d\tau}\hat{c}&=-\frac{i}{\hbar}[\hat{c},\hat{H}]\\&=\exp(i\delta_q\tau)\kappa_q^*\{\varphi_1^2(\tau)(\hat{I}-2\hat{c}^\dagger\hat{c})\\&\quad\times\exp[i(2\lambda-\delta)\tau]-2\varphi_1(\tau)\varphi_2^*(\tau)\hat{c}\}\hat{a}_q\\&\quad+\exp(-i\delta_q\tau)\kappa_q^*\{\varphi_2^2(\tau)(\hat{I}-2\hat{c}^\dagger\hat{c})\\&\quad\times\exp[i(2\lambda-\delta)\tau]-2\varphi_1^*(\tau)\varphi_2(\tau)\hat{c}\}\hat{a}_q^\dagger,\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{d}{d\tau}\hat{a}_q&=-\frac{i}{\hbar}[\hat{a}_q,\hat{H}]\\&=\exp(-i\delta_q\tau)\kappa_q\{\varphi_1^*(\tau)\varphi_2(\tau)(\hat{I}-2\hat{c}^\dagger\hat{c})\\&\quad+\exp[i(2\lambda-\delta)\tau]\varphi_2^2(\tau)\hat{c}^\dagger\\&\quad-\exp[-i(2\lambda-\delta)\tau]\varphi_1^{*2}(\tau)\hat{c}\},\end{aligned} \tag{12}$$

where $\delta_q\equiv(\Omega-\omega_q)/\omega'$ is the difference in the frequencies of the classical and quantized fields, normalized to the modulation frequency, $\kappa_q=\kappa\exp(-i\psi)$, and $\hat{a}_q=\hat{a}\exp(i\omega_q t)$. When analyzing the solutions of the system of equations (11), we assume that the quantized field is weak, i.e., $|\kappa_q|=|\kappa|\ll 1$. Note that spectroscopically the classical field is strong, i.e., the coupling constant σ is of the order of unity or higher. This makes it possible to ignore the dynamics of the classical field and interpret it as a fixed external disturbance.

As noted earlier, the functions $\varphi_1(\tau)$ and $\varphi_2(\tau)$ are 2π-periodic. In Ref. 23 it was shown that the problem of the interaction of an atom and a classical field can always be formulated in such a way that $\varphi_1(\tau)$ is a linear combination of only even harmonics and $\varphi_2(\tau)$, a linear combination of only odd harmonics:

$$\begin{aligned}\varphi_1(\tau)&=\sum_{n=-\infty}^{\infty}a_n\exp(-2in\tau),\\\varphi_2(\tau)&=\sum_{m=-\infty}^{\infty}b_m\exp[i(2m+1)\tau].\end{aligned} \tag{13}$$

Equations (11)–(13) and the above assumption that the parameter κ is small imply that the interaction of the ''dressed'' atom and the field is of a resonant nature [the right-hand sides of Eqs. (11) and (12) acquire slowly oscillating terms] only if one of the following approximate equalities holds:

$$\begin{aligned}&(1)\ \ 2\lambda-\delta+\delta_q\approx 2m_1\ \text{ or }\ \omega_q\approx\omega_1-2\omega' m_1,\\&(2)\ \ 2\lambda-\delta-\delta_q\approx 2m_2\ \text{ or }\ \omega_q\approx\omega_2+2\omega' m_2,\\&(3)\ \ \delta_q\approx 2m_3+1\ \text{ or }\ \omega_q\approx\omega_3-2\omega' m_3,\end{aligned} \tag{14}$$

where $\omega_1=\omega+2\lambda\omega'$, $\omega_2=\omega+2(\delta-\lambda)\omega'$, and $\omega_3=\omega+(\delta-1)\omega'$ are the transition frequencies between the quasienergy levels shown in Fig. 1, and m_1, m_2, and m_3 are arbitrary integers that number the quasienergy levels in the equidistant sequences $0'$, $0''$, $1'$, and $1''$.

In cases (1) and (2), the quantum transitions are between quasienergy levels, which correspond to different quasienergy states. In case (1) the lower state is $|\theta_0\rangle$ ($|\theta_1\rangle$) and the upper state is $|\theta_1\rangle$ ($|\theta_0\rangle$). This type of interaction can be interpreted as ''inelastic.''

In case (3), the upper and lower quasilevels correspond to the same quasienergy state, $|\theta_0\rangle$ or $|\theta_1\rangle$. Thus, despite the interaction with the quantized mode, the quasienergy states $|\theta_0\rangle$ and $|\theta_1\rangle$ remain the stationary states of the atom. This type of interaction is a feature specific to open systems and can be interpreted as ''elastic.'' In an ''elastic'' interaction the atom emits a photon into the quantized mode or absorbs a photon, but the quasienergy of the atom remains unchanged (within our approach the quasienergy is defined modulo $2\hbar\omega'$).

Thus, what mechanism of atom–field interaction, ''inelastic'' or ''elastic,'' is realized depends on the frequency ω_q of the quantized mode. The difference between these two mechanisms disappears if the quasilevels belonging to different sequences move closer to each other within a distance of the order of $\hbar\omega'$ (parametric resonance; see Refs. 20, 25, and 28). Below we assume that the minimum distance be-

tween the quasienergy levels exceeds $\hbar\omega'$ considerably, which makes it possible to study the ''inelastic'' and ''elastic'' mechanisms separately.

In ''inelastic'' interaction [case (1)], the frequency of the quantized mode is given by the relation

$$\delta_q=\delta-2\lambda+2m_1+\varepsilon, \quad |\varepsilon|\ll 1, \tag{15}$$

where m_1 is the order of the resonance, and ε is the normalized detuning of the frequency ω_q of the quantized mode from the resonance frequency $\omega_1-2\omega' m_1$. Substituting (15) into the right sides of Eqs. (11) and (12) and dropping all terms except the slowly oscillating, we obtain

$$\frac{d}{d\tau}\hat{c}=\kappa_{m_1}^*\exp(i\varepsilon\tau)(\hat{I}-2\hat{c}^\dagger\hat{c})\hat{a}_q,$$

$$\frac{d}{d\tau}\hat{a}_q=-\kappa_{m_1}\exp(-i\varepsilon\tau)\hat{c}, \tag{16}$$

where

$$\kappa_{m_1}=\kappa_q\langle\varphi_1^{*2}(\tau)\exp(-2im_1\tau)\rangle \tag{17}$$

is the effective atom–field coupling constant, with the angle brackets standing for averaging over the period 2π.

Transforming the photon annihilation operator according to the expression $\hat{a}_q=\exp(-i\varepsilon\tau)\hat{a}_1$, we can write Eq. (16) in the form of a Heisenberg equation:

$$\frac{d}{d\tau}\hat{c}=-\frac{i}{\hbar}[\hat{c},\hat{H}_{m_1}^{\text{eff}}], \quad \frac{d}{d\tau}\hat{a}_1=-\frac{i}{\hbar}[\hat{a}_1,\hat{H}_{m_1}^{\text{eff}}], \tag{18}$$

where the effective Hamiltonian is defined as

$$\hat{H}_{m_1}^{\text{eff}}=-\hbar\omega'\varepsilon\hat{a}_1^\dagger\hat{a}_1-i\hbar\omega'(\kappa_{m_1}\hat{a}^\dagger\hat{c}-\kappa_{m_1}^*\hat{a}\hat{c}^\dagger). \tag{19}$$

Thus, in the ''inelastic'' interaction approximation, our model is formally equivalent to the initial Jaynes–Cummings model.[1] The parameters of the model are the detuning from resonance, ε, and the coupling constant κ_{m_1}. By way of an example, let us examine the case where the atom is placed into a bichromatic (amplitude-modulated) field, whose median frequency coincides with the transition frequency ω:

$$\Omega=\omega, \quad \delta=0, \quad g(\tau)=\cos\tau. \tag{20}$$

The quasienergy states are defined in (3). Here[28]

$$\lambda=0, \quad \varphi_1(\tau)=\cos(\sigma\sin\tau), \quad \varphi_2(\tau)=\sin(\sigma\sin\tau), \tag{21}$$

and the expression (17) for the coupling constant becomes

$$\kappa_{m_1}=\begin{cases}\dfrac{1}{2}\kappa_q[1+J_0(2\sigma)]=\dfrac{1}{2}\kappa\exp(-i\psi)[1+J_0(2\sigma)],\\ m_1=0,\\ \dfrac{1}{2}\kappa_q J_{2m_1}(2\sigma)=\dfrac{1}{2}\kappa\exp(-i\psi)J_{2m_1}(2\sigma),\\ m_1\neq 0.\end{cases} \tag{22}$$

Thus, the observation of higher-order ''inelastic'' resonances ($m_1\gg 1$) is possible only in strong classical fields ($\sigma\gg 1$).

The second case of ''inelastic'' interaction,

$$\delta_q=2\lambda-\delta-2m_2-\varepsilon, \quad |\varepsilon|\ll 1, \tag{23}$$

can be examined in the same way. The fact that the upper and lower states in the quantum transition have traded places leads to an interchange in the operators $\hat{c}$ and $\hat{c}^\dagger$ in the expression for the effective Hamiltonian:

$$\hat{H}_{m_2}^{\text{eff}}=\hbar\omega'\varepsilon\hat{a}_2^\dagger\hat{a}_2+i\hbar\omega'(\kappa_{m_2}a_2^\dagger c^\dagger-\kappa_{m_2}^*a_2 c), \tag{24}$$

where

$$\hat{a}_q=\exp(i\varepsilon\tau)\hat{a}_2, \quad \kappa_{m_2}=\kappa_q\langle\varphi_2^2(\tau)\exp(2im_2\tau)\rangle. \tag{25}$$

Let us examine the ''elastic'' interaction:

$$\delta_q=2m_3+1+\varepsilon, \quad |\varepsilon|\ll 1. \tag{26}$$

Keeping only the slowly oscillating terms on the right sides of Eqs. (11) and (12), we find

$$\frac{d}{d\tau}\hat{c}=2\kappa_{m_3}^*\exp(i\varepsilon\tau)\hat{c}\hat{a}_q-2\kappa_{m_3}\exp(-i\varepsilon\tau)\hat{c}\hat{a}_q^\dagger, \tag{27}$$

$$\frac{d}{d\tau}\hat{a}_q=\kappa_{m_3}\exp(-i\varepsilon\tau)(\hat{I}-2\hat{c}^\dagger\hat{c}), \tag{28}$$

where

$$\kappa_{m_3}=\kappa_q\langle\varphi_1^*(\tau)\varphi_2(\tau)\exp[-i(2m_3+1)\tau]\rangle. \tag{29}$$

Transforming the photon annihilation operator according to the expression $\hat{a}_q=\exp(i\varepsilon\tau)\hat{a}_3$, we can write Eqs. (27) and (28) in the form of Heisenberg equations with an effective Hamiltonian

$$H_{m_3}^{\text{eff}}=-\hbar\omega'[\varepsilon\hat{a}_3^\dagger\hat{a}_3+i(\kappa_{m_3}^*\hat{a}_3-\kappa_{m_3}\hat{a}_3^\dagger)(\hat{I}-2\hat{c}^\dagger\hat{c})]. \tag{30}$$

The operator $\hat{N}(\tau)=\hat{c}^\dagger(\tau)\hat{c}(\tau)$ commutes with the Hamiltonian (30) and is a constant of motion of the system:

$$\hat{c}^\dagger(\tau)\hat{c}(\tau)=\hat{c}^\dagger(0)\hat{c}(0)=\hat{c}_0^\dagger\hat{c}_0. \tag{31}$$

We see that when the interaction is ''elastic,'' populations of the quasienergy states $|\theta_0\rangle$ and $|\theta_1\rangle$ do not vary with time. Substituting (31) into (28) and integrating the equation for the photon annihilation operator, we obtain

$$\hat{a}_q(\tau)=\hat{a}_q(0)+i\kappa_{m_3}\frac{\exp(-i\varepsilon\tau)-1}{\varepsilon}(\hat{I}-2\hat{c}_0^\dagger\hat{c}_0), \tag{32}$$

$$\begin{aligned}\hat{n}(\tau)=\hat{a}^\dagger(\tau)\hat{a}(\tau)=\hat{a}_q^\dagger(\tau)\hat{a}_q(\tau)=\hat{n}_0\\ +4|k_{m_3}|^2\frac{\sin^2(\varepsilon\tau/2)}{\varepsilon^2}+\frac{i}{\varepsilon}(\hat{I}-2\hat{c}_0^\dagger\hat{c}_0)\\ \times\{\kappa_{m_3}[\exp(-i\varepsilon\tau)-1]a_0^\dagger-\kappa_{m_3}^*\\ \times[\exp(i\varepsilon\tau)-1]a_0\},\end{aligned} \tag{33}$$

where $\hat{a}_0\equiv\hat{a}(0)$, and $\hat{n}_0\equiv\hat{a}_0^\dagger\hat{a}_0$. In deriving relation (33) for the operator $\hat{n}(\tau)$ of the photons in the quantized mode, we employed the fact that $(\hat{I}-2\hat{c}_0^\dagger\hat{c}_0)^2=\hat{I}$. When the interaction is ''elastic,'' the oscillations of the photons take place at the

frequency $\varepsilon\omega'$. In exact resonance between the ''dressed'' atom and the quantized mode ($\varepsilon=0$), the number of photons is

$$\hat{n}(\tau)=\hat{n}_0+|\kappa_{m_3}|^2\tau^2+(\hat{I}-2\hat{c}_0^{\dagger}\hat{c}_0)(\kappa_{m_3}a_0^{\dagger}-\kappa_{m_3}^{*}a_0)\tau. \tag{34}$$

We see that, in contrast to ''inelastic'' interaction, ''elastic'' interaction makes it possible to effectively pump energy from a classical field to a quantized mode. The reason for this is that for the ''elastic'' interaction there is no saturation of the population inversion of the quasienergy states.

Let us assume that initially the ''dressed'' atom and the quantized mode were statistically independent, and that the quantized mode was in a coherent state $|\alpha\rangle$. Using Eq. (33), we can derive the following expressions for the average number of photons, $\hat{n}$, and the variance of the number of photons, (Δn^2):

$$\overline{n(\tau)}=|\alpha|^2+|\eta(\tau)|^2+\gamma(\tau)\Delta N, \tag{35}$$

$$\frac{(\Delta n)^2}{\overline{n}}=1+\frac{\gamma(\tau)^2[1-(\Delta N)^2]}{\overline{n}}, \tag{36}$$

where $\Delta N=\langle\hat{I}-2\hat{c}_0^{\dagger}\hat{c}_0\rangle=\langle\hat{c}_0\hat{c}_0^{\dagger}-\hat{c}_0^{\dagger}\hat{c}_0\rangle$ is the average population difference of the quasienergy states $|\theta_0\rangle$ and $|\theta_1\rangle$ (the angle brackets denote quantum mechanical averaging), and

$$\eta(\tau)=\frac{i\kappa_{m_3}[\exp(-i\varepsilon\tau)-1]}{\varepsilon},$$

$$\gamma(\tau)=\alpha^*\eta(\tau)+\alpha\eta^*(\tau).$$

Thus, at $\alpha=0$, the photon statistics in the quantized mode is represented by a Poisson distribution at any moment in time $[(\Delta n)^2=\overline{n}]$. The same is true for an arbitrary value of α if the atom is in one of its quasienergy states, i.e., $\Delta N=\pm 1$. In all other cases the photon statistic is above-Poisson $[(\Delta n)^2>\overline{n}]$.

To estimate the highest order of the resonance, m_3, that can be realized in an experiment we must average over the period 2π in (29). In particular, for the amplitude-modulated (bichromatic) classical field (20) we have

$$\kappa_{m_3}=-\frac{i}{2}\kappa_q J_{2m_3+1}(2\sigma)$$

$$=-\frac{i}{2}\kappa\,\exp(-i\psi)J_{2m_3+1}(2\sigma). \tag{37}$$

4. CONCLUSIONS

We have studied the nonlinear dynamics of a coupled system which incorporates a ''dressed'' atom and one quantized mode of an electromagnetic field. By ''dressed'' we mean a two-level atom that interacts with an intensive classical field. The approximations which we used are those of a dipole interaction, a rotating wave, and a fixed classical field.

If the classical field has an equidistant frequency spectrum, two stationary quasienergy states and an infinite number of quasienergy levels can be introduced for the ''dressed'' atom. The interaction of the ''dressed'' atom and the quantized mode is of a resonant nature if the frequency of the quantized mode is close to one of the transition frequencies between the quasienergy levels. What atom–field interaction mechanism, ''inelastic'' or ''elastic'' (which differ considerably), is realized depends on the frequency of the quantized mode.

In the ''inelastic'' mechanism, the quasienergy levels between which a quantum transition takes place belong to different quasienergy states of the ''dressed'' atom. Here emission or absorption of a photon from the quantized mode is accompanied by a transition of the ''dressed'' atom from one quasienergy state to the other. The equations of ''inelastic'' interaction can formally be reduced to equations describing the dynamics of the initial Jaynes–Cummings model (a two-level atom and a single quantized mode).

In the ''elastic'' mechanism, both quasienergy levels belong to the same quasienergy state. Emission or absorption of a photon is not accompanied by a change in the populations of the quasienergy states. When the frequency of the quantized mode and that of the transition between quasienergy levels coincide, there can be an infinite buildup of the number of photon in the quantized mode with the passage of time (as long as the approximation of a fixed classical field holds). If initially the quantized mode was in a coherent quantum state $|\alpha\rangle$, with the passage of time the statistics of the photons in the quantized mode either becomes of the Poisson type $[(\Delta n)^2=\overline{n}]$ or of the above-Poisson type $[(\Delta n)^2>\overline{n}]$. The case of Poisson statistics is realized either at $\alpha=0$ or at an arbitrary value of α if the atom is in one of its quasienergy states.

[1] E. T. Jaynes and F. W. Cummings, Proc. IEEE **51**, 87 (1963).
[2] J. R. Ackerharlt and K. Rzazewski, Phys. Rev. A **12**, 2549 (1975).
[3] N. B. Narozhny, J. J. Sanchez-Mondragon, and J. H. Eberly, Phys. Rev. A **22**, 236 (1981).
[4] L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms*, Wiley, New York (1975).
[5] J. H. Eberly, N. B. Narozhny, and J. J. Sanchez-Mondragon, Phys. Rev. Lett. **44**, 1323 (1980).
[6] P. Goldberg and L. C. Harrison, Phys. Rev. A **43**, 376 (1991).
[7] H.-I. Yoo and J. H. Eberly, Phys. Rep. **118**, 241 (1985).
[8] H.-S. Li and N.-Y. Bei, Phys. Lett. A **101**, 169 (1984).
[9] N. N. Bogolubov, Jr., F. L. Klein, and A. S. Shumovski, Phys. Lett. A **101**, 201 (1984).
[10] V. V. Dodonov, V. I. Man'ko, and S. M. Chumakov, in *Proc. (Trudy) of the P. N. Lebedev Physics Institute* [in Russian], Vol. **176, Moscow (1986), p.** 57.
[11] J. L. Gruver, J. Aliaga, H. A. Cerdeira, and A. N. Proto, Phys. Rev. A **50**, 5274 (1994).
[12] D. Meschede, H. Walther, and G. Muller, Phys. Rev. Lett. **54**, 551 (1985).
[13] G. Rempe, H. Walther, and N. Klein, Phys. Rev. Lett. **58**, 353 (1987).
[14] M. M. Ashraf, Phys. Rev. A **50**, 5116 (1994).
[15] H. Huang and H. Fan, Phys. Lett. A **166**, 308 (1992).
[16] H. Huang and H. Fan, Phys. Lett. A **159**, 323 (1991).
[17] A. Ya. Kazakov, Phys. Lett. A **206**, 229 (1995).
[18] A. Ya. Kazakov, Opt. Spektrosk. **81**, 549 (1996) [Opt. Spectrosc. **81**, 498 (1996)].
[19] H. T. Dung and A. S. Shumovski, Phys. Lett. A **169**, 379 (1992).
[20] C. Cohen-Tannoudji, in *Cargese Lectures in Physics*, Vol. 2, M. Levy (ed.), Gordon and Breach, New York (1968), p. 347
[21] T. Yabuzaki, S. Nakayama, Y. Marakami, and T. Ogawa, Adv. At. Mol. Phys. **10**, 1955 (1974).
[22] J. H. Shirley, Phys. Rev. **138**, B979 (1965).

[23] M. Z. Smirnov, Kvant. Élektron. (Moscow) **22**, 903 (1995).
[24] M. Z. Smirnov, J. Opt. Soc. Am. B **9**, 2191 (1992).
[25] M. Z. Smirnov, J. Opt. Soc. Am. B **11**, 109 (1994).
[26] Y. Zhu, Q. Wu, A. Lezama, D. J. Gauthier, and T. W. Mossberg, Phys. Rev. A **41**, 6574 (1990).
[27] G. S. Agarwal, Y. Zhu, D. J. Gauthier, and T. W. Mossberg, J. Opt. Soc. Am. B **8**, 1163 (1991).
[28] M. Z. Smirnov, Phys. Rev. A **52**, 2195 (1995).

Translated by Eugene Yankovsky
Edited by S. J. Amoretty

Critical fields for ionization of the hydrogen molecule and the molecular hydrogen ion

M. B. Smirnov and V. P. Kraĭnov

Moscow Physics and Technology Institute, 141700 Dolgoprudnyĭ, Moscow Region, Russia
(Submitted 13 March 1997)
Zh. Éksp. Teor. Fiz. **112**, 828–835 (September 1997)

We calculate the critical fields for the classical ionization of an ion of the hydrogen molecule and the neutral hydrogen molecule by an electric field. In the case of a molecular ion we examine different internuclear distances and obtain the correct limits for the well-known cases of small and large internuclear distances. © *1997 American Institute of Physics.* [S1063-7761(97)00509-X]

1. INTRODUCTION

Bethe and Salpeter[1] were the first to obtain the value of the critical electric field strength needed to ionize atoms according to the classical mechanics:

$$F=\frac{E_n^2}{4Z}\ \text{a.u.} \tag{1}$$

Here F is the electric field strength at which the unperturbed energy E_n of the atomic level under consideration coincides with the top of the effective potential barrier, and Z is the charge of the atomic (or ionic) core. (Throughout the paper we use the atomic system of units: $e=\hbar=m_e=1$.) In particular, for a hydrogenlike state with the principal quantum number n, Eq. (1) yields

$$F=\frac{Z^3}{16n^4}\ \text{a.u.} \tag{2}$$

One should bear in mind, however, that the estimates (1) and (2) were obtained for the one-dimensional model. In other words, it is assumed that the electron moves along the direction of the electric field. The three-dimensional case changes somewhat only the values of the numerical coefficients in (1) and (2).

If the electric field strength exceeds the values (1) or (2), classical above-barrier ionization of the atomic occurs over a very short atomic time. In the quantum case, the ionization probability proves to be somewhat lower because of above-barrier reflection. When the field is weak, there can only be quantum tunneling through the effective potential barrier, with the tunneling rate being exponentially small compared to the classical ionization rate.

Actually, in Eq. (1) we must take the perturbed value of the energy of the level under consideration, i.e., with allowance for the Stark shift. But if, for example, we take the ground state of the hydrogen atom, even for the critical value of the electric field the Stark shift amounts to

$$-\frac{9F^2}{4}=-\frac{9(1/16)^2}{4}\approx-0.01\ \text{a.u.},$$

which is extremely small compared to the unperturbed energy value $E=-0.50$ a.u. Thus, the Stark shift in (1) can be ignored. The same is true of the ground states of the simplest molecules. In any case, the one-dimensional approximation is much cruder. An exception is the molecular ion of hydrogen, since an external electric field strongly mixes the ground even-parity state and the first excited odd-parity state, which results in a large Stark shift of these levels with respect to each other at large internuclear distances.

The goal of the present study is to obtain the values of the critical electric fields for the simplest molecules, primarily for the ionization of an ion of the hydrogen molecule and the neutral hydrogen molecule. This problem is important because of the recent experiments in the interaction of high-power low-frequency laser light and simple diatomic molecules and their molecular ions.[2,3] While for atoms the critical fields can be obtained analytically, for molecules we are forced to employ numerical methods, since the values of these fields depend on the distance between the nuclei of the diatomic molecule or the molecular ion. What is more, for a molecular hydrogen ion the Stark shifts of the electron levels are much larger than in the case of atoms because of the large induced dipole moments. As for atoms, we use the one-dimensional approximation to simplify the problem. Thus, the results are of semiquantitative accuracy. A systematic coverage of the results of the numerous experimental and theoretical studies of the molecular hydrogen ion (as the simplest molecular entity) by laser methods can be found in the review of Giusti-Suzor *et al.*[4]

The estimates (1) and (2) are true for a constant electric field and a low-frequency laser field, but in the latter case the alternating perturbation must be adiabatic, i.e.,

$$\gamma=\frac{\omega\sqrt{2E_n}}{F}\ll 1.$$

For low-lying states of atoms and molecules and for fields obeying estimates (1) and (2), this condition is satisfied to a good accuracy for the radiation of a carbon-dioxide laser.

2. THEORETICAL TREATMENT OF THE MOLECULAR HYDROGEN ION

Let us denote the distance between the protons in the ion of a hydrogen molecule by R and the coordinate of an electron parallel to the external electric (quasistationary) field of strength F by z. The effective potential with which the protons and the external field act on this electron has the following simple form:

$$V(z)=-\frac{1}{|z-R/2|}-\frac{1}{|z+R/2|}-Fz. \tag{3}$$

Qualitatively, the chief difference between the effective potential of a molecule and that of an atom is that while in the atom an electron surmounts only one barrier, in the case of a molecule or molecular ion there are two potential barriers: one to the left and the other to the right of the origin of coordinates (in agreement with Eq. (3)). Hence, in classical ionization the electron must surmount a higher potential barrier, and depending on the energy of the excited level being considered, field strength, and internuclear distance, the higher barrier may be either to the left or to the right of the origin. All this complicates the calculations.

It is well known (see, e.g., the review in Ref. 5 and Slater's monograph[6]) that the interaction of a molecular hydrogen ion and an external constant or low-frequency electric field can be treated in the approximation of the two lowest electron states: the ground (even-parity) state with an energy $E_g(R)$ and the first excited (odd-parity) state with an energy $E_u(R)$. The values of these energies as functions of the internuclear distance R have been thoroughly studied.[7] For convenience we change the signs of these energies, thus assuming that both are positive.

In an external electric field the even- and odd-parity states strongly shift in relation to each other because of a finite dipole moment that couples these states. Hence the requirement that the energy of the perturbed state be equal to the height of the effective potential barrier amounts to the following:

$$V(z)=E_{g,u}(F,R), \quad \frac{dV}{dz}=0. \tag{4}$$

The perturbed energies $E_{g,u}(F,R)$ can be determined from the following equation:[5]

$$\det\begin{vmatrix} E_{g,u}(F,R)+E_g(R) & FR/2 \\ FR/2 & E_{g,u}(F,R)+E_u(R) \end{vmatrix}=0. \tag{5}$$

Here we have assumed that the dipole matrix element between the even- and odd-parity states under consideration is equal to $R/2$. Strictly speaking, this is true only when the distance R between the protons is large. However, numerical calculations have demonstrated that this approximation can also be used in the range of equilibrium distances, $R=2.0$ a.u. (Ref. 8). Thus, we arrive at the following simple expressions for the perturbed energies:

$$E_g(F,R)=-\frac{E_g(R)+E_u(R)}{2} - \sqrt{\left(\frac{E_g(R)-E_u(R)}{2}\right)^2+\left(\frac{FR}{2}\right)^2}, \tag{6}$$

$$E_u(F,R)=-\frac{E_g(R)+E_u(R)}{2} + \sqrt{\left(\frac{E_g(R)-E_u(R)}{2}\right)^2+\left(\frac{FR}{2}\right)^2}. \tag{7}$$

Let us start with the ground state. Substituting (6) and (7) in (4) and eliminating F from the two equations in (4), we arrive at an equation for the coordinate z at which this state touches the top of the effective potential barrier:

$$8z^3=[E_g(R)+E_u(R)]\left(z^2-\frac{R_2}{4}\right)^2 + \sqrt{[E_g(R)-E_u(R)]^2\left(z^2-\frac{R^2}{4}\right)^4+4R^2\left(z^2+\frac{R^2}{4}\right)^2}. \tag{8}$$

Then the critical electric field strength is given by the following expression:

$$F(R)=2\,\frac{z^2+R^2/4}{(z^2-R^2/4)^2}\,. \tag{9}$$

Note that here $z>R/2$, so that the value of the perturbed energy $E_g(F,R)$ is equal to the top of the right potential barrier.

In the $R\to 0$ limit we find $E_g\to 2$ a.u., $z=2$ a.u., and $F=0.5$ a.u., which agrees with (2) for $Z=2$ and $n=1$ (the ground state of the He^+ ion). In the opposite limit $R\to\infty$ we have $E_g=E_u=0.5$ a.u. and $z=R/2+4+8/R+\cdots$ a.u., so that $F=1/6$ a.u., which agrees with (2) for $Z=n=1$ (the ground state of the hydrogen atom).

Now let us calculate the critical electric field strength for the first excited (odd-parity) state of the molecular hydrogen ion. The calculation is more complicated than in the case of the ground (even-parity) state. If the distance R between the nuclei is not very large, the values of the critical field $F(R)$ can be found from the condition that the energy of the perturbed odd-parity state coincide with the top of the right potential barrier, since this barrier is higher than the left potential barrier. The corresponding coordinate z can be found numerically from an equation obtained in the same way as Eq. (8) (only the sign in front of the root is reversed):

$$8z^3=[E_g(R)+E_u(R)]\left(z^2-\frac{R^2}{4}\right)^2 - \sqrt{[E_g(R)-E_u(R)]^2\left(z^2-\frac{R^2}{4}\right)^4+4R^2\left(z^2+\frac{R^2}{4}\right)^2}. \tag{10}$$

The critical electric field strength F is also found from Eq. (9) (just as it is for the even-parity state) by reasoning along similar lines. In particular, in the $R\to 0$ limit we obtain $E_u\to 0.5$ a.u., so that $z=8$ a.u. and $F=1/32$ a.u. This agrees with (2) if $Z=2$ and $n=2$ (the $2p$ state of the helium atom).

When the internuclear distance R is large, the critical field strength is found from the condition that the perturbed energy of the odd-parity state coincide with the top of the left potential barrier, which becomes higher than the right barrier. Here we have $-R/2<z<0$. (The results of numerical calculations suggest that the left and right potential barriers are of equal height at $R=7.5$ a.u.) In this case the effective potential barrier has the following form:

$$V(z)=\frac{1}{z-R/2}-\frac{1}{z+R/2}-Fz. \tag{11}$$

Then the conditions

$$V(z)=E_u(F,R), \quad \frac{dV}{dz}=0 \tag{12}$$

TABLE I. The critical electric field strengths F_g and F_u and the corresponding coordinates z_g and z_u as functions of the proton–proton distance R for a molecular hydrogen ion (all values are given in the atomic system of units).

R	$E_g(R)$	$E_u(R)$	z_g	z_u	F_g	F_u
0	2.000	0.500	2.00	8.00	0.500	0.0312
0.5	1.735	0.521	2.35	7.70	0.375	0.0338
1.0	1.445	0.583	2.90	6.94	0.260	0.0422
1.5	1.249	0.625	3.44	6.59	0.195	0.0479
2.0	1.102	0.667	3.96	6.37	0.155	0.0531
2.5	0.994	0.687	4.44	6.42	0.129	0.0544
3.0	0.904	0.693	4.89	6.66	0.111	0.0526
3.5	0.847	0.693	5.24	6.94	0.102	0.0504
4.0	0.796	0.692	5.59	7.33	0.095	0.0467
4.5	0.756	0.684	5.91	7.74	0.090	0.0432
5.0	0.724	0.675	6.22	8.16	0.085	0.0400
5.5	0.699	0.667	6.51	8.58	0.082	0.0372
6.0	0.678	0.656	6.80	9.00	0.080	0.0347
6.5	0.661	0.647	7.07	9.40	0.078	0.0326
7.0	0.648	0.647	7.35	9.80	0.076	0.0308
7.5	0.636	0.630	7.63	10.2	0.074	0.0290
8.0	0.627	0.623	7.90	−0.52	0.073	0.0336
8.5	0.618	0.616	8.16	−0.69	0.072	0.0379
9.0	0.612	0.610	8.42	−0.87	0.071	0.0412
9.5	0.605	0.605	8.68	−1.06	0.070	0.0438
10.0	0.510	0.510	54.0	−46.0	0.0625	0.0625

yield an equation for the coordinate $z<0$ at which the energy of the perturbed odd-parity state coincides with the top of the left potential barrier:

$$2R\left(3z^2-\frac{R^2}{4}\right)=-[E_g(R)+E_u(R)]\left(z^2-\frac{R^2}{4}\right)^2 + \sqrt{[E_g(R)-E_u(R)]^2\left(z^2-\frac{R^2}{4}\right)^4+4R^4z^2}. \quad (13)$$

The critical electric field strength in this case is given by the following expression:

$$F=-\frac{2zR}{(z^2-R^2/4)^2}. \quad (14)$$

In the limit $R\to\infty$ we have $E_g=E_u=0.5$ a.u., and from (13) and (14) it follows that $z=-R/2+4$ a.u. and that $F=1/16$ a.u., as it should for the ground state of the hydrogen atom, according to (2).

3. RESULTS AND DISCUSSION OF THE CASE OF THE H_2^+ ION

Table I lists the results of calculations of the critical field strength F for the even- and odd-parity states of the molecular hydrogen ion as functions of the internuclear distance R. We also list the values of the coordinate z at which the energy of the corresponding state coincides with the top of the effective potential barrier. Finally, we list the unperturbed values of the energies of the even- and odd-parity electron states used in the calculations.

We see that the critical field strength for the odd-parity state is much lower than for the even-parity state, as expected (with, of course, the exception of the case of extremely large internuclear distances, where the two values coincide). Note the change of sign of the contact coordinate z for the odd-parity state at the point where the right barrier is replaced by the left one (at $R=7.5$ a.u.).

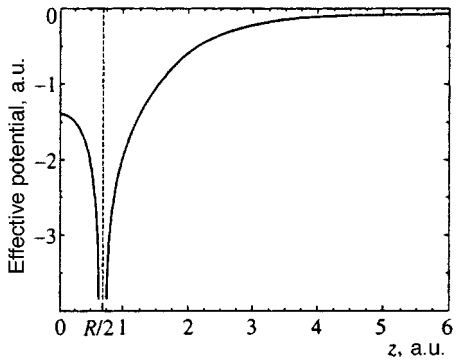

FIG. 1.

Table I shows that the critical field strength F_u for the ground (even-parity) state of the molecular hydrogen ion monotonically decreases from the value 0.5 a.u. at $R=0$, corresponding to the $1s$ state of the He^+ ion, to the value $1/16=0.0625$ a.u. as $R\to\infty$, corresponding to the $1s$ state of the hydrogen atom (as expected). The behavior of the critical field strength for the odd-parity state as a function of the internuclear distance R is more complicated: at small values of R the critical field strength grows from the value $1/32=0.0312$ a.u. at $R=0$, corresponding to the $2p$ state of the He^+ ion, to a local maximum at $R=2.5$ a.u. After that F_u decreases to a local minimum at $R=7.5$ a.u., and then again increases to the final value $1/16=0.0625$ a.u. as $R\to\infty$, corresponding to the $1s$ state of the hydrogen atom.

The minimum value of F_u at $R=7.5$ a.u. (corresponding to a radiation intensity $I=5\times10^{13}$ W cm^{-2}) explains the maximum electron ionization probability at such an internuclear distance; indeed, an electron cannot be ''ionized'' from the ground electron state because of the presence of a potential barrier; instead it is nonadiabatically transferred from the ground electron state to an excited state because of the Landau–Zener effect[9,10] and is then ''ionized'' from the excited state. Note that the data listed in Table I agree with the results of calculations of Posthumus *et al.*[11] of the ionization probability (see Fig. 1 in Ref. 11).

4. CRITICAL FIELD STRENGTH FOR THE NEUTRAL HYDROGEN MOLECULE

Calculations of the critical field strength for the neutral hydrogen molecule differ considerably from those of the molecular hydrogen ion. In the neutral molecule, in contrast to the ion, an external electric field does not mix the ground state with an excited electron state of the molecule, since the dipole matrix element does not increase with the internuclear distance. We can therefore ignore the Stark shift of the energy of the ground state of the neutral hydrogen molecule. This conclusion is confirmed by the results of Yu *et al.*[12] In addition, ionization by a low-frequency laser field occurs at the equilibrium internuclear distance $R_e=1.458$ a.u., since there is practically no dissociation in the neutral hydrogen

molecule. This is confirmed by experiments.[2,3] Hence there is no need to consider other values of the internuclear distance.

The variational three-dimensional wave function of the ground state of the hydrogen molecule is well known:[6]

$$\Psi(\mathbf{r}_1,\mathbf{r}_2)=\exp[-\alpha(r_a(1)+r_b(2))]+\exp[-\alpha(r_a(2)+r_b(1))] \tag{15}$$

(its normalization is not required in the present problem). Here we have introduced the following notation: $\alpha=1.166$ a.u., and $r_{a,b}(1,2)$ are the distances between the electrons *1* and *2* and the protons *a* and *b*:

$$\begin{aligned} r_a(1)&=\sqrt{\rho_1^2+(z_1+R/2)^2},\\ r_b(1)&=\sqrt{\rho_1^2+(z_1-R/2)^2},\\ r_a(2)&=\sqrt{\rho_2^2+(z_2+R/2)^2},\\ r_b(2)&=\sqrt{\rho_2^2+(z_2-R/2)^2}. \end{aligned} \tag{16}$$

The electron coordinates $\mathbf{r}_1=(\rho_1,z_1)$ and $\mathbf{r}_2=(\rho_2,z_2)$ are expressed in terms of cylindrical coordinates, with the axis coinciding with the direction of the external electric field.

Next we define the single-particle potential of one of the electrons (say, the first) in the field of the second electron by using the Hartree–Fock approximation:

$$V_e(\mathbf{r}_1)=\frac{\int \frac{1}{|\mathbf{r}_1-\mathbf{r}_2|}|\Psi(\mathbf{r}_1,\mathbf{r}_2)|^2 d\mathbf{r}_2}{\int|\Psi(\mathbf{r}_1,\mathbf{r}_2)|^2 d\mathbf{r}_2}. \tag{17}$$

Then the effective potential acting on the first electron in the process of its ''ionization'' has the following form (here we again assume that the electron is emitted along the direction of the external electric field, i.e., we employ the one-dimensional approximation):

$$V(z)=-\frac{1}{|z-R/2|}-\frac{1}{|z+R/2|}+V_e(\rho=0,z)-Fz. \tag{18}$$

The effective single-particle potential corresponding to (18) at $F=0$ is depicted in Fig. 1 (it is symmetric with respect to the origin of the coordinate system).

The critical field strength F and the corresponding coordinate z is found (as in the case of the molecular hydrogen ion) by numerically solving the following two equations:

$$\frac{dV}{dz}=0,\quad V(z)=\frac{1}{2}E, \tag{19}$$

where $E=-1.139$ a.u. is the energy of the ground state of the hydrogen molecule. Here we also assume that the two electrons are identical, i.e., each has an average energy equal to one-half of the total energy.

A simple calculation that uses (18) and (19) yields

$$F=0.0694 \text{ a.u.},\quad z=4.43 \text{ a.u.} \tag{20}$$

We see that the critical field strength is close to 1/16 a.u. for the ground state of the hydrogen atom. This can be explained by the screening of the Coulomb field of the two protons by the field of the second electron.

Thus, we have calculated the critical electric field strengths for the case of classical ionization of the simplest molecules: the molecular hydrogen ion and the neutral molecule of hydrogen. They supplement the results for atoms.[13] Our results can be used to analyze the process of simultaneous dissociation and ionization of a molecular hydrogen ion initialized by high-power low-frequency laser radiation (primarily of a carbon-dioxide laser with the maximum amplitudes of the field strength equaling 10^8 V cm^{-1} or higher; see Refs. 14 and 15).

The authors are grateful to S. P. Goreslavskiĭ, N. B. Delone, A. M. Popov, and M. V. Popov for discussing the results of the research and for helpful remarks. The work was sponsored by the Russian Fund for Fundamental Research (Grants Nos. 95-02-03657 and 96-02-18299).

[1] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Systems*, Springer, Berlin (1958).
[2] F. A. Ilkov, T. D. G. Walsh, S. Turgeon, and S. L. Chin, Phys. Rev. A **51**, R2695 (1995).
[3] F. A. Ilkov, T. D. G. Walsh, S. Turgeon, and S. L. Chin, Chem. Phys. Lett. **247**, 1 (1995).
[4] A. Giusti-Suzor, F. H. Mies, L. F. DiMauro, E. Charron, and B. Yang, J. Phys. B **28**, 309 (1995).
[5] P. Dietrich and P. B. Corkum, J. Chem. Phys. **97**, 3187 (1992).
[6] J. C. Slater, *Quantum Theory of Molecules and Solids*, Vol. 1: *Electron Structure of Molecules*, McGraw-Hill, New York (1963).
[7] D. R. Bates, K. Ledsham, and A. L. Stewart, Philos. Trans. R. Soc. London, Ser. A **246**, 215 (1953).
[8] M. E. Sukharev and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz **110**, 832 (1996) [JETP **83**, 457 (1996)].
[9] T. Seideman, M. Yu. Ivanov, and P. B. Corkum, Phys. Rev. Lett. **75**, 2819 (1995).
[10] P. Dietrich, M. Yu. Ivanov, F. A. Ilkov, and P. B. Corkum, Phys. Rev. Lett. **77**, 4150 (1996).
[11] J. H. Posthumus, M. R. Thompson, L. J. Frasinski, and K. Codling, Phys. Rev. Lett. **77**, 298 (1996).
[12] H. Yu, T. Zuo, A. D. Bandrauk, Phys. Rev. A **54**, 3290 (1996).
[13] N. B. Delone and V. P. Krainov, *Multiphoton Processes in Atoms*, Springer, Berlin (1994).
[14] M. Ivanov, T. Seideman, P. Corkum, F. Ilkov, and P. Dietrich, Phys. Rev. A **54**, 1541 (1996).
[15] Z. Milyukov, M. Pont, and R. Shakeshaft, Phys. Rev. A **54**, 4299 (1966).

Translated by Eugene Yankovsky

Nonlinear saturation spectroscopy of the degenerate electron gas in spherical metallic particles

S. G. Rautian

Institute of Automation and Electrometry, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia
(Submitted 21 March 1997)
Zh. Éksp. Teor. Fiz. **112**, 836–855 (September 1997)

We examine the linear and cubic optical polarizations of the degenerate electron gas in spherical metallic particles. For photon energies that are high compared to the splitting of the spectrum of electron states we use a unified approach to calculate the dimensional (size, or quantum) and Drude parts of the polarization and the variation of energy distribution of the electrons induced by a field. © *1997 American Institute of Physics.* [S1063-7761(97)00609-4]

1. INTRODUCTION

Spherical metallic particles play an important role in many physical problems, with the result that over the years their various properties—mechanical, thermal, magnetic, electrical, radiophysical, optical—have undergone study. In particular, the optical properties, which we investigate here, have been studied for more than 90 years, starting with the pioneering work of Maxwell–Garnett[1] and Mie.[2] In the 1950s there was an upsurge of interest in metal optics in connection in the hope that it could be used to solve important problems of the physics of metals.[3] The interest in quantum (size) effects in microparticles much larger than the interatomic distances in condensed media had an impact on the physics of metallic particles. In the 1960s, Kabawata and Kubo[4] and Gor'kov and Éliashberg[5] studied the specific heat and magnetic and electric properties of small particles caused by quantum effects. The nonlinear optical phenomena exhibited in a high-power laser field are also influenced by size effects, and according to Hache *et al.*,[6] these effects provide the leading contribution to the optical nonlinear susceptibility tensor of the electrons in spherical metallic particles. Unfortunately, we doubt that this is the case, and the reason for this is given below.

The theory of size optical effects in microparticles proved to be closely linked to the general problem of selecting the Hamiltonian $\hat{V}$ of the interaction of an electron and an electromagnetic field, a problem that has been constantly debated over the entire period in which nonlinear optics and spectroscopy developed.[7–9] The Hamiltonian $\hat{V}$ can be written in the form

$$\hat{V}=-\frac{e}{mc}\,\hat{\mathbf{p}}\cdot\mathbf{A}+\frac{e^2}{2mc^2}\mathbf{A}^2, \tag{1.1}$$

where $\hat{\mathbf{p}}$ and $\mathbf{A}$ are the electron momentum and the field's vector potential. In the dipole approximation, Eq. (1.1) is equivalent to[7–9]

$$\hat{V}=-\hat{\mathbf{d}}\cdot\mathbf{E}, \tag{1.2}$$

where $\hat{\mathbf{d}}$ and $\mathbf{E}$ are the dipole moment and the electric field vector. The equivalence, however, breaks down if in the perturbation series we keep only a limited number of terms, which describe the mixing of only some unperturbed electron states. The reason is that the matrix elements of the Hamiltonians (1.1) and (1.2) for the transitions between stationary states m and n differ by a factor ω_{mn}/ω, where ω_{mn} and ω are the Bohr frequency and the frequency of the field. In resonance ω_{mn} equals ω, and both Hamiltonians obviously lead to the same result. From general considerations it is clear that for quantum effects in particles of nonatomic sizes the situation is markedly different. Indeed, the characteristic scale in the energy spectrum is the quantity

$$E_0=\frac{\hbar^2}{2ma^2}=R\left(\frac{a_0}{a}\right)^2, \quad R=\frac{me^4}{2\hbar^2}=1.1\times 10^5 \ \text{cm}^{-1}, \tag{1.3}$$

where a is the particle radius, and R and a_0 are the Rydberg constant and the Bohr radius. If, for instance, $a=100a_0=53$ Å, then $E=10\ \text{cm}^{-1}\ll\hbar\omega=10^4\ \text{cm}^{-1}$. Thus, for a sizable fraction of the quantum transitions in microparticles of nonatomic sizes we have $|\omega_{mn}|\ll\omega$, and for such transitions the matrix elements of the Hamiltonians (1.1) and (1.2) differ by several orders of magnitude, and the choice between (1.1) and (1.2) becomes critical. Analysis of a number on nonlinear phenomena has shown that for approximate calculations the Hamiltonian (1.2) is preferable to (1.1) (see Refs. 8 and 10). But the ''preferability criteria'' formulated in Refs. 8 and 10 work only for the simplest model (two- and three-level). In our problem of the interaction of the degenerate electron gas and a field the model is essentially multilevel: the number of filled states is equal to the number of electrons in a particle and amounts to roughly one million for a particle with a radius of 100 Å. We see that the criteria developed in Refs. 8 and 10 are of little use here. Below we will show that in calculating the spectrum of nonlinear absorption of electrons in spherical particles we should still use formula (1.2), with the results differing both in absolute value and in the dependence on the field frequency from those obtained in the ''alternative theory'' of Hache *et al.*,[6] based on the Hamiltonian (1.1).

In this paper we develop a theory of optical (linear and nonlinear) properties of the degenerate electron gas in spherical particles. We employ a simple model, in which an electron is in an infinitely deep spherical potential well of radius a. The electron wave function in this case is[11]

$$\psi_{nlm}=\frac{C_{nl}Y_{lm}(\theta,\varphi)J_{l+1/2}(\alpha_{nl}r/a)}{\sqrt{r}}, \tag{1.4}$$

$$C_{nl}=\frac{\sqrt{2}}{a|J_{l+3/2}(\alpha_{nl})|},$$

where $Y_{lm}(\theta,\varphi)$ and $J_{l+1/2}(z)$ are the spherical function and the Bessel function of the first kind, α_{nl} is the $(n+1)$st root of $J_{l+1/2}(z)$, and C_{nl} is a normalization constant. The energy eigenvalues E_{nl} of the state nl are

$$E_{nl}=E_0\alpha_{nl}^2, \quad E_0=\frac{\hbar^2}{2ma^2}. \tag{1.5}$$

Some of the properties of the roots of Bessel functions that we will need in the future can be found in Appendix A.

For typical values $a=100$ Å we have $E_0=3.08\ \mathrm{cm}^{-1}$, and for a Fermi energy $E_F=4.5\times10^4\ \mathrm{cm}^{-1}$ (silver) the maximum value $\alpha_{nl}=\alpha_F$ for filled states (at low electron temperatures) amounts to $\alpha_F=\sqrt{E_F/E_0}=120$. Thus, we are dealing with large quantum numbers: in the given numerical example $n\leqslant 38$ and $l\leqslant 110$. The characteristic separation of excited levels,

$$E_i-E_j=E_0(\alpha_i^2-\alpha_j^2)=\sqrt{E_0}(\sqrt{E_i}+\sqrt{E_j})(\alpha_i-\alpha_j), \tag{1.6}$$

$$i=n_il_i, \quad j=n_jl_j,$$

for $\alpha_i-\alpha_j=1$, and $\alpha_i\approx a_j\approx\alpha_F$ amounts to 740 cm^{-1}, which is much larger than the relaxation widths ($\sim100\ \mathrm{cm}^{-1}$) but much smaller than $\hbar\omega$.

The off-diagonal matrix element of the radius, r_{ij}, is given by the following expression:

$$r_{ij}=\frac{4a\alpha_i\alpha_j}{(\alpha_i^2-\alpha_j^2)^2}(\delta_{l_il_j-1}+\delta_{l_il_j+1}). \tag{1.7}$$

We see that we have $r_{ij}\leqslant a$, with r_{ij} assuming its maximum value at the minimum difference $|\alpha_i-\alpha_j|=1$. This fact is similar to the well-known rule of atomic spectroscopy, according to which the lines are the strongest for equal values of the principal quantum numbers of the combining states.[12]

The model (1.4) is certainly a simplification. First, it does not account for the roughness of the particle surface. Since the electron wavelength is $\lambda=\hbar/\sqrt{2mE_F}=a/\alpha_F\sim10^{-2}a$, irregularities in the shape of the surface with sizes on the order of interatomic distances may become important. Second, the single-particle approximation ignores the electron–electron interaction, which, however, is characteristic of the concept of a degenerate electron gas. The concept of a quantum Fermi liquid may also improve the theory. Nevertheless, the simple free-electron model (1.4) has proved useful in many physical problems. For our purposes it is simply indispensable since, we believe that incorporating the above factors into the theory of nonlinear spectroscopic properties of particles constitutes an exceptional and, perhaps, unjustified complication of the model.

2. THE DIPOLE MOMENT OF A SPHERE: GENERAL EXPRESSIONS

Since we assume that the sphere is small ($a\ll\lambda$), we can limit ourselves to the dipole approximation and define the dipole moment as follows:

$$\langle\mathbf{d}(t)\rangle=\mathrm{Tr}\ \hat{\mathbf{d}}\rho=\sum_{ij}\ \mathbf{d}_{ji}\rho_{ij}, \tag{2.1}$$

where $\hat{\mathbf{d}}$ and ρ is the dipole moment operator and the density matrix. In the model of relaxation constants, the kinetic equation for ρ has the form (see, e.g., Refs. 13 and 14)

$$\dot{\rho}_{ij}+(\Gamma_{ij}+i\omega_{ij})\rho_{ij}=2i\sum_l\ (G_{il}\rho_{lj}-\rho_{il}G_{lj})\times\cos\ \omega t+\delta_{ij}\Gamma_jN_j, \tag{2.2}$$

where ω is the field frequency, the Γ_{ij} are relaxation constants, and the N_j are the populations of the states $j=n_jl_jM_j$ in the absence of a field. Here we have introduced the following notation:

$$\Gamma_j=\Gamma_{jj}, \quad \omega_{ij}=\frac{E_i-E_j}{\hbar}, \tag{2.3}$$

$$G_{\alpha\beta}=\frac{\mathbf{d}_{\alpha\beta}\cdot\mathbf{E}}{2\hbar}=\sum_\sigma\ \frac{d_{\alpha\beta\sigma}(-1)^\sigma E_{-\sigma}}{2\hbar}, \tag{2.4}$$

$$\mathbf{e}_\sigma\cdot\mathbf{d}_{\alpha\beta}=d_{\alpha\beta\sigma}=\frac{d_{\alpha\beta}(-1)^{l_\beta-M_\beta}}{\sqrt{3}}\langle l_\alpha M_\alpha l_\beta-M_\beta|1\sigma\rangle, \tag{2.5}$$

$$d_{\alpha\beta}=(n_\alpha l_\alpha\|d\|n_\beta l_\beta)=\frac{4ea\alpha_\alpha\alpha_\beta\sqrt{l_{\max}^{\alpha\beta}}\delta_{l_\alpha l_\beta\pm1}}{(\alpha_\alpha^2-\alpha_\beta^2)^2},$$

$$l_{\max}^{\alpha\beta}=\max\ l_\alpha l_\beta, \tag{2.6}$$

with $d_{\alpha\beta\sigma}$ and E_σ the circular components of the vector $d_{\alpha\beta}$ and of the field amplitude E, where $d_{\alpha\beta}$ is a simplified notation for the reduced matrix element of the dipole moment defined in accordance of Ref. 12, and $\langle\cdots|\cdots\rangle$ is a vector addition coefficient. The labels i, j, α, β indicate sets of quantum numbers $n_il_iM_i$, $n_jl_jM_j$, etc. The operator representing the interaction with the field is written in the form dE, in accordance with what has been said in Sec. 1.

It is common practice in nonlinear optics to use Cartesian coordinates for the susceptibility tensor, which is quite natural for crystalline media. In our problem the object is spherically symmetric, and so we use a more convenient tool, irreducible spherical tensors, which has found wide application in nuclear physics,[15] atomic spectroscopy,[12] the theory of relaxation processes,[16] and in nonlinear spectroscopy.[13,14]

Relaxation processes are represented in Eq. (2.2) by phenomenological constants, which describe the relaxation of levels (Γ_j) and coherence (Γ_{ij}). Such a relaxation model is widely used, although it has certain drawbacks.

The solution of Eq. (2.2) can be written in the form of a Fourier series:

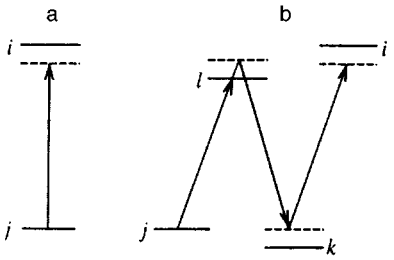

FIG. 1. Diagrams for (a) one-photon and (b) three-photon transitions.

$$\rho_{ij}=\sum_{s=-\infty}^{\infty}\rho_{ij}^{(s)}\exp(i\omega st),\quad \rho_{ij}^{(s)}=(\rho_{ji}^{(-s)})^*. \tag{2.7}$$

The component of the dipole moment oscillating with a frequency ω is

$$\mathbf{d}_\omega=\sum_{ij}\mathbf{d}_{ji}(\rho_{ij}^{(-1)}e^{-i\omega t}+\rho_{ij}^{(1)}e^{i\omega t})=\mathrm{Re}(\mathbf{D}e^{-i\omega t}),\quad \mathbf{D}=2\sum_{ij}\mathbf{d}_{ji}\rho_{ij}^{(-1)}. \tag{2.8}$$

The expansion amplitudes $\rho_{ij}^{(s)}$ in (2.7) obey the following system of equations:

$$[\Gamma_{ij}+i(\omega_{ij}+s\omega)]\rho_{ij}^{(s)}=\delta_{s0}\delta_{ij}\Gamma_i N_j+i\sum_l[G_{il}(\rho_{lj}^{(s-1)}+\rho_{lj}^{(s+1)})-(\rho_{il}^{(s-1)}+\rho_{il}^{(s+1)})G_{lj}]. \tag{2.9}$$

The solution $\rho_{ij}^{(-1)}$ needed for calculating $\mathbf{d}_\omega$ can be written, to within terms cubic in the field amplitude, as follows:

$$\rho_{ij}^{(-1)}=G_{ij}L_{ij}^{(-1)}N_{ji}+\sum_{kl}G_{ik}G_{kl}G_{lj}B_{iklj}, \tag{2.10}$$

where

$$L_{\alpha\beta}^{(s)}=\frac{1}{\omega_{\alpha\beta}+s\omega-i\Gamma_{\alpha\beta}},\quad N_{\alpha\beta}=N_\alpha-N_\beta, \tag{2.11}$$

$$B_{iklj}=L_{ij}^{(-1)}\{[(L_{kj}^{(0)}+L_{il}^{(0)})(L_{kl}^{(-1)}+L_{kl}^{(1)})+(L_{kj}^{(-2)}+L_{kl}^{(-2)})L_{kl}^{(-1)}]N_{lk}+[L_{kj}^{(0)}(L_{lj}^{(-1)}+L_{lj}^{(1)})+L_{kj}^{(-2)}L_{lj}^{(-1)}]N_{jl}+[L_{il}^{(0)}(L_{ik}^{(-1)}+L_{ik}^{(1)})+L_{il}^{(-2)}L_{ik}^{(-1)}]N_{ki}\}. \tag{2.12}$$

Thus, the formula for the circular component D_σ is

$$D_\sigma=\mathbf{e}_\sigma\cdot\mathbf{D}=2\sum_{ij}d_{ij\sigma}G_{ij}L_{ij}^{(-1)}N_{ji}+2\sum_{iklj}d_{ji\sigma}G_{ik}G_{kl}G_{lj}B_{iklj}. \tag{2.13}$$

The term in (2.13) that is linear in the field corresponds to the one-photon transition $j-i$ (Fig. 1a), while the nonlinear (cubic) term corresponds to the three-photon transition $j-l-k-i$ (Fig. 1b), in accordance with the ordinary nonlinear spectroscopic scheme of saturation of atoms, molecules, and, in general, systems with a discrete spectrum. At each stage in Figs. 1a and b, the resonance conditions specified by the coefficients $L_{\alpha\beta}^{(s)}$ are important. The quantities $L_{\alpha\beta}^{(\pm1)}$ describe the Lorentzian shapes of the lines for the allowed transitions $\alpha-\beta$. The factors $L_{\alpha\beta}^{(0)}$ are related to the constant terms $\rho_{\alpha\beta}^{(0)}$, with the transition $\alpha-\beta$ being forbidden in this case. The coefficients $L_{\alpha\beta}^{(-2)}$ represent beats (with a frequency 2ω) of populations ($\alpha=b$) and matrix elements $\rho_{\alpha\beta}$ for forbidden transitions ($j-k$ and $l-i$). In earlier work these beats were ignored.

The distribution of the populations N_j over the states plays an important role in finding the sums in (2.13). We assume that the N_j are independent of the M_j. In addition, all calculations are done for low electron temperatures, where the Fermi distribution can be assumed rectangular. For high-power radiation this condition is not met, but here this fact will be ignored. In accordance with what has just been said,[1)]

$$N_j=0,\quad E_j>E_F;\quad N_j=1,\quad E_j<E_F. \tag{2.14}$$

Summation over the magnetic sublevels in (2.13) can be done by standard methods: only the vector addition coefficients in $d_{ji\sigma}$ and $G_{\alpha\beta}$ depend on the magnetic numbers, and summation of their produces is done according to well-known rules (see, e.g., Refs. 12–14 and 17). The result is

$$D_\sigma=\frac{1}{3\hbar}\sum_{ij}|d_{ij}|^2L_{ij}^{(-1)}N_{ji}E_\sigma+\frac{I_0}{12\hbar^3}\sum_{jlki}d_{ji}^*d_{ik}d_{lk}^*d_{lj}\times B_{iklj}\sum_{\sigma_1\kappa q}A_{jlki}^{(\kappa)}\langle 1\sigma_1\kappa q|1\sigma\rangle E_{\sigma_1}I(\kappa q), \tag{2.15}$$

$$I(\kappa q)=\sqrt{3}\sum_{\sigma_2\sigma_3}(-1)^{1-\sigma_3}\langle 1\sigma_2 1-\sigma_3|\kappa q\rangle\frac{E_{\sigma_2}E_{\sigma_3}^*}{I_0},\quad I_0=\sum_\sigma|E_\sigma|^2, \tag{2.16}$$

$$\frac{2A_{jlki}^{(\kappa)}}{\sqrt{2\kappa+1}}=(-1)^{l_j-l_k}\begin{Bmatrix}1&1&\kappa\\ l_j&l_k&l_i\end{Bmatrix}\begin{Bmatrix}1&1&\kappa\\ l_j&l_k&l_l\end{Bmatrix}+(-1)^{l_l-l_i}\begin{Bmatrix}1&1&\kappa\\ l_l&l_i&l_j\end{Bmatrix}\begin{Bmatrix}1&1&\kappa\\ l_l&l_i&l_k\end{Bmatrix}. \tag{2.17}$$

We have retained the previous notation for the summation indices in (2.15), the only difference being that now they stand for the sets n_jl_j, n_il_i, etc. The quantities denoted in Eq. (2.17) by braces are $6j$-symbols.[12,17,18] The quantity $I(\kappa q)$ is known as the polarization tensor of the field,[13,14] and the chosen normalization of the tensor $I(\kappa q)$ is such that $I(00)=1$; we isolate the factor I_0 proportional to the intensity of the radiation. The tensor $I(\kappa q)$, the $A_{jlki}^{(\kappa)}$, and the vector addition coefficient in (2.15) describe the anisotropic properties of the cubic nonlinearity. For instance, for linearly polarized radiation (with the z axis directed along E) we have

$$I(00)=1,\quad I(20)=-\sqrt{2}, \tag{2.18}$$

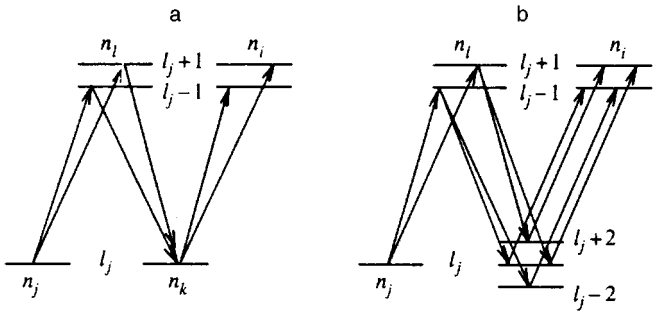

FIG. 2. Diagrams illustrating the change in the orbital quantum number in three-photon transitions for $\kappa=0$ and 1 (a) and for $\kappa=2$ (b).

while for circular polarization (with the quantization axis directed along the wave vector)

$$I(00)=1, \quad I(10)=\sqrt{\frac{3}{2}}, \quad I(20)=\frac{1}{\sqrt{2}}. \tag{2.19}$$

The other components of $I(\kappa q)$ vanish.

The expression for the cubic component D_σ in (2.15) possesses a certain ''hidden'' symmetry, which is quite evident in the initial expression (2.13): the sum over $M_jM_iM_kM_i$ and over the indices of the circular components of the field retains its value under the following substitutions of the indices: $j\to l\to k\to i\to j$ and $\sigma_1\to\sigma_2\to\sigma_3\to\sigma_1$. At the same time, each product of $6j$-symbols in (2.17) is unsymmetric under such substitutions, but the sums over $\sigma_1\kappa q$ incorporating these products are the same. Hence the $A^{(\kappa)}_{jlki}$ could have been written in the form of any of the two products of $6j$-symbols. In Eq. (2.17) we used a symmetrized notation, in which the above-mentioned symmetry is incorporated explicitly into the expression for the coefficients $A^{(\kappa)}_{jlki}$.

According to the properties of $6j$-symbols,[12,17,18] the $A^{(\kappa)}_{jlki}$ are finite if

$$|l_j-l_l|\leqslant 1, \quad |l_l-l_k|\leqslant 1, \quad |l_k-l_i|\leqslant 1, \quad |l_i-l_j|\leqslant 1, \tag{2.20}$$

$$|l_j-l_k|\leqslant\kappa, \quad |l_l-l_i|\leqslant\kappa. \tag{2.21}$$

In view of the selection rules for r_{ij}, only equalities are realized in (2.20). And due to the same selection rules, at $\kappa=0$ and 1 we have $l_j=l_k$ and $l_l=l_i$, and only at $\kappa=2$ are the inequalities in (2.21) also realized. The selection rules in the orbital quantum number are illustrated by the transition diagrams in Fig. 2.

Since the statistical weight of an l state is equal to $2l+1$, large values of orbital quantum numbers are important. Calculations have shown that for $l\gg 1$ the coefficients $A^{(\kappa)}_{jlki}$ are inversely proportional to l_j. For instance, for the isotropic part we have

$$A^{(0)}_{jlki}=\frac{1}{6l_j}, \quad l_j\gg 1. \tag{2.22}$$

3. THE TWO-LEVEL APPROXIMATIONS

Summation over the quantum numbers in (2.15) is somewhat different in the linear and cubic terms of the dipole moment. For this reason we introduce the special notation

$$D_\sigma=D^{(1)}_\sigma+D^{(3)}_\sigma. \tag{3.1}$$

We begin with the linear component $D^{(1)}_\sigma$. In the sum over ij in (2.15), the indices ij run through the same sets of values. We select a pair of levels mn and assume $E_m>E_n$. The indices ij appear in the sum in two combinations:

$$i=m, \quad j=n; \quad i=n, \quad j=m. \tag{3.2}$$

Taking into account their total contribution to $D^{(1)}_\sigma$, we find that

$$D^{(1)}_\sigma=\frac{2}{3\hbar}\sum_{mn}\frac{\omega_{mn}|d_{mn}|^2N_{nm}}{\omega^2_{mn}-(\omega+i\Gamma_{mn})^2}E_\sigma, \tag{3.3}$$

where summation over the $m-n$ transitions is implied. The sum in the cubic term $D^{(3)}_\sigma$ is generally much more complicated than in the linear. However, many terms in the sum over $jlki$ are nonresonant and provide a small contribution to D_σ. Following Ref. 6, we adopt the two-level approximation, in which

$$k=j, \quad l=i. \tag{3.4}$$

Passing to summation over transitions, as we did in the linear case, we arrive at the following expression:

$$\begin{aligned}D^{(3)}_\sigma=&-\frac{I_0}{3\hbar^3}\sum_{mn}\frac{\omega_{mn}|d_{mn}|^4N_{nm}}{\omega^2_{mn}-(\omega+i\Gamma_{mn})^2}\left\{\left(\frac{1}{\Gamma_m}+\frac{1}{\Gamma_n}\right)\right.\\&\times\left[\frac{\Gamma_{mn}}{(\omega_{mn}-\omega)^2+\Gamma^2_{mn}}+\frac{\Gamma_{mn}}{(\omega_{mn}+\omega)^2+\Gamma^2_{mn}}\right]\\&\left.+\left(\frac{1}{2\omega+i\Gamma_m}+\frac{1}{2\omega+i\Gamma_n}\right)\frac{\omega+i\Gamma_{mn}}{\omega^2_{mn}-(\omega+i\Gamma_{mn})^2}\right\}\\&\times\sum_{\sigma_1\kappa q}\langle 1\sigma_1\kappa q|1\sigma\rangle E_{\sigma_1}I(\kappa q)A_{mn\kappa},\end{aligned} \tag{3.5}$$

$$A_{mn\kappa}=\frac{1}{2}\sqrt{2\kappa+1}\left[\begin{Bmatrix}1&1&\kappa\\l_m&l_m&l_n\end{Bmatrix}^2+\begin{Bmatrix}1&1&\kappa\\l_n&l_n&l_m\end{Bmatrix}^2\right]. \tag{3.6}$$

Summation over the scalar quantum numbers n_ml_m and n_nl_n in Eqs. (3.3) and (3.5) has nontrivial aspects specific to the model (1.4). We illustrate these aspects using the simpler expression (3.3) for $D^{(1)}_\sigma$. In what follows we assume that the relaxation constants Γ_{mn} and Γ_m are the same for all levels and transitions:

$$\Gamma_m=\Gamma_n=\Gamma_1, \quad \Gamma_{mn}=\Gamma_2. \tag{3.7}$$

According to (2.6), $d_{mn}\propto\omega^{-2}_{mn}$, which means that there are two regions of Bohr frequencies ω_{mn} contributing the most to $D^{(1)}_\sigma$ in the sum in (3.3): the region of small ω_{mn} and the region near the point $\omega_{mn}=\omega$. If

$$\frac{2\sqrt{E_0E_F}}{\hbar\omega}\ll 1, \tag{3.8}$$

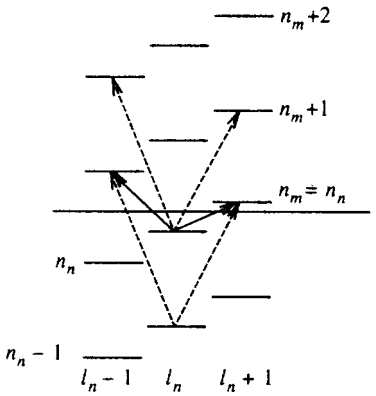

FIG. 3. Diagram representing the nonresonant transitions $n_n, l_n \to n_n + k, l_n + 1$ (right-slanted arrows) and $n_n, l_n \to n_n + 1 + k, l_n - 1$ (left-slanted arrows). Solid arrows correspond to $k=0$ and dashed arrows to $k=1$. Transitions with $k \geqslant 2$ are not shown.

which we assume to be the case, these regions do not overlap and summation in the two can be done independently.[2)] The inequality (3.8) imposes a restriction on the ratio a/λ, which in a certain sense is the opposite of the criterion for the applicability of the dipole approximation; specifically, the inequality (3.8) can be written as

$$\frac{a}{\lambda} \gg \frac{v_F}{2\pi c} \approx 10^{-3}, \tag{3.8a}$$

where v_F is the electron velocity on the Fermi surface (for silver $v_F = 1.4 \times 10^8$ cm s^{-1}). Because v_F/c is small, the condition (3.8) does not contradict the inequality $a/\lambda \ll 1$, although it does limit the interval within which the particle size can vary.

We start with small values of ω_{mn} and denote the corresponding component of $D_\sigma^{(1)}$ by $D_{\sigma D}^{(1)}$. In the denominators of the terms of the series we drop the term ω_{mn}^2 and insert the expression for d_{mn} from formula (2.6):

$$D_{\sigma D}^{(1)} = -\frac{2E_\sigma}{3\hbar} \frac{1}{(\omega + i\Gamma_2)^2} \Sigma, \tag{3.9}$$

$$\Sigma = \sum_{mn} \omega_{mn} |d_{mn}|^2 N_{nm}$$

$$= (4ae)^2 \frac{E_0}{\hbar} \sum_{mn} \frac{\alpha_m^2 \alpha_n^2 l_{\max}^{mn} N_{nm}}{(\alpha_m^2 - \alpha_n^2)^3} \delta_{l_m l_n \pm 1}.$$

At low electron temperatures the difference $N_{nm} = N_n - N_m$ is finite for $E_n < E_F$ and $E_m > E_F$, i.e., the energy levels m and n are close to the Fermi surface but are on different sides of it (Fig. 3). In the numerator we put

$$\alpha_m = \alpha_n = \alpha_F = \sqrt{\frac{E_F}{E_0}},$$

in the denominator we put

$$\alpha_m^2 - \alpha_n^2 = 2\alpha_F(\alpha_m - \alpha_n),$$

then sum over l_m, and use the interpolation formula (A4) for the roots α_m and α_n:

$$\Sigma = \frac{e^2}{m} \hbar \alpha_F \sum_{l_n n_n n_m} \left\{ \frac{l_n + 1}{[Q(n_m - n_n) + P]^3} + \frac{l_n}{[Q(n_m - n_n - 1) + Q - P]^3} \right\}. \tag{3.10}$$

The first term in the braces describes $l_n \to l_m = l_n + 1$ transitions (in Fig. 3). Clearly, $n_m - n_n = k = 0, 1, 2, \ldots$ for such transitions, and for a fixed k there is a $(k+1)$st transition of this type (the right-slanted arrows; in Fig. 3 only the transitions with $k=0$ and $k=1$ are shown). The $l_n \to l_m = l_n - 1$ transitions are described by the second term in the braces. For such transitions the minimum value is $n_m - n_n = 1$, i.e., $n_m - n_n - 1 = k = 0, 1, 2, \ldots$, and for a fixed k there is a $(k+1)$st transition of this type (the left-slanted arrows in Fig. 3). Thus,

$$\Sigma = \frac{e^2}{m} \hbar \alpha_F \sum_{l_n} \sum_{k=0}^{\infty} \left[\frac{(l_n+1)(k+1)}{(Qk+P)^3} + \frac{l_n(k+1)}{(Qk+Q-P)^3} \right]. \tag{3.11}$$

The series in k in (3.11) can be expressed by a linear combination of generalized Riemann zeta functions defined by the following equality:[19,20]

$$\zeta(s,v) = \sum_{k=0}^{\infty} (k+v)^{-s}. \tag{3.12}$$

The functions $\zeta(s,v)$ have been tabulated (see, e.g., Ref. 21), but for our analysis of the dependence on physical parameters a simple approximate expression is more convenient (see Appendix B):

$$\zeta(s,v) = \frac{1}{v^s} \left[1 + \frac{v^s}{(s-1)(v+1/2)^{s-1}} \right], \tag{3.13}$$

which we will use in our investigation. Summation over l_n in (3.11) now amounts to averaging in the interval $0 < l_n < l_{\max} \approx \alpha_F$ over a layer of thickness $\alpha_m - \alpha_n = \Delta\alpha = P$ or $Q - P$ adjacent to the Fermi surface. The number of states N whose energy does not exceed E_n is[22]

$$N = \frac{V}{3\pi^2} \left(\frac{2m}{\hbar^2} E_F \right)^{3/2} = \frac{4}{9\pi} \left(\frac{E_F}{E_0} \right)^{3/2} = \frac{4}{9\pi} \alpha_F^3, \tag{3.14}$$

where $V = (4\pi/3)a^3$. Hence the number of state in a layer of thickness $\Delta\alpha$ is

$$\Delta N = \frac{4}{3\pi} \alpha_F^2 \Delta\alpha. \tag{3.15}$$

Collecting the intermediate results, we arrive at the following relationship:

$$D_{\sigma D}^{(1)} = -\frac{e^2 N}{m(\omega + i\Gamma_2)^2} F_1, \tag{3.16}$$

where

$$F_1 = 2 \left\langle \frac{l_n+1}{P^2} \left[1 + \frac{2P^3(P+2Q)}{Q^2(2P+Q)^2} \right] + \frac{l_n}{(Q-P)^2} \times \left[1 + \frac{2(Q-P)^3(3Q-P)}{Q^2(3Q-2P)^2} \right] \right\rangle [\langle 2l_n + 1 \rangle]^{-1}, \tag{3.17}$$

with the angle brackets denoting summation along the surface $\alpha_n=\alpha_F$ in a layer of unit thickness, $\Delta\alpha=1$. The factor F_1 cannot be calculated analytically, but its value is close to unity. Indeed, for $l_n\ll n_n$ we have $Q=2P=\pi$, and calculations via formula (3.17) yield $F_1=37/4\pi^2=0.94$. In the opposite limit $l_n\approx\alpha_F\gg 1$ we get $P-1\ll 1$ and $Q>\pi$ (see Appendix A). At $Q=5$ and $P=1$, for instance, we have $F_1=1.10$. Thus, we may assume that $F_1=1$.

In calculating the nonlinear component of the dipole moment, $D_{\sigma D}^{(3)}$ we must allow for two important facts. First, the factor $A_{mn\kappa}$ in (3.6) can be replaced by its value for $l_j\gg 1$:

$$A_{mn\kappa}=\frac{A_\kappa}{6l_j},\quad A_0=1,\quad A_1=\frac{\sqrt{3}}{2},\quad A_2=\frac{\sqrt{5}}{10}. \tag{3.18}$$

Second, the terms in the series in mn in (3.5) contain the factor $(\alpha_m^2-\alpha_n^2)^{-7}$, and with such a large exponent we need only keep the terms with $n_m=n_n$ for $l_m=l_n+1$ and with $n_m=n_n+1$ for $l_m=l_n-1$. In the rest, the calculation can be done according to the above scheme, which leads to the following relationship:

$$D_{\sigma D}^{(3)}=\frac{e^2N}{m(\omega+i\Gamma_2)^2}\left(\frac{ea}{\hbar\omega}\right)^2 I_0\left\{\frac{\Gamma_2}{\Gamma_1}-\frac{1}{4(1+i\Gamma_1/2\omega)(1+i\Gamma_2/\omega)}\right\}C_\sigma F_3, \tag{3.19}$$

where

$$C_\sigma=\frac{1}{3}\sum_{\sigma_1\kappa q}\langle 1\sigma_1\kappa q|1\sigma\rangle E_{\sigma_1}I(\kappa q)A_\kappa, \tag{3.20}$$

$$F_3=\langle 2l_n[P^{-6}+(Q-P)^{-6}]\rangle\,[\langle 2l_n+1\rangle]^{-1}. \tag{3.21}$$

The angle brackets in (3.21) imply, as usual, summation over a layer of unit thickness, $\Delta\alpha=1$, near the Fermi surface.

If in (3.16) we put $F_1=1$, then $D_{\sigma D}^{(1)}$ is equal to the dipole moment calculated in the classical Drude free-electron model. This result is quite natural since $D_{\sigma D}^{(1)}$ is determined by the nonresonant interaction of the field and an electron with the Fermi energy, as assumed in the classical theory of optical properties of metals.[3] Hence the term $D_{\sigma D}^{(3)}$ can be called the Drude nonlinearity, which is indicated by the subscript D. The nonlinear parameter in $D_{\sigma D}^{(3)}$ is the combination $(ea/\hbar\omega)^2I_0$, which stresses the nonlinearity of the interaction; the characteristic buildup time of the nonlinear effect is $1/\omega$.

The term in (3.19) containing $(1+i\Gamma_1/2\omega)^{-1}$ reflects the presence of population beats. This effect, not considered in Refs. 4–6, is indeed unimportant if $\Gamma_2\gg\Gamma_1$.

Now let us calculate the contribution to D_σ from the region of states near the point $\omega_{mn}=\omega$. This is called the quantum or dimensional (size) or resonant contribution, since it is caused by the resonant transitions between states of the discrete spectrum related to the finite dimensions (or size) of the particles. We denote it by $D_{\sigma a}$.

Summation in (3.3) and (3.5) over l_m yields $2l_n+1$. We replace the sum over n_m by an integral and use the result to calculate the contribution of the resonant regions, e.g.,

$$\sum_{n_m}\frac{1}{\omega_{mn}^2-(\omega+i\Gamma_2)^2}=\int_0^\infty\frac{d\omega_{mn}}{\omega_{mn}^2-(\omega+i\Gamma_2)^2}\frac{dn_m}{d\omega_{mn}}=i\frac{\pi}{2}\frac{1}{\omega+i\Gamma_2}\frac{dn_m}{d\omega_{mn}},$$

$$\frac{d\omega_{mn}}{dn_m}=\frac{2Q}{\hbar}\sqrt{E_0E_m}, \tag{3.22}$$

where the coefficient Q is weakly dependent on n. In all other factors we set $\omega_{mn}=\omega$, $\alpha_m^2=\alpha_n^2+\hbar\omega/E_0$, etc. The sum over n_nl_n is replaced by an integral over E_n with a density of states

$$\frac{dN}{dE_n}=\frac{2}{3\pi}\frac{E_n^{1/2}}{E_0^{3/2}}.$$

As a result of the calculations we arrive at the following relationships:

$$D_{\sigma a}^{(1)}=i\frac{e^2N}{m}\frac{1}{\omega(\omega+i\Gamma_2)}\frac{2\sqrt{E_0E_F}}{\hbar\omega}g_1(\nu)E_\sigma, \tag{3.23}$$

$$g_1(\nu)=\frac{1}{\nu}\int_{1-\nu}^1\chi^{3/2}(\chi+\nu)^{1/2}\frac{\pi}{Q}d\chi,\quad \nu=\frac{\hbar\omega}{E_F}, \tag{3.24}$$

$$D_{\sigma a}^{(3)}=-i\frac{e^2N}{m\omega^2}\left(\frac{ea}{\hbar}\right)^2\frac{I_0}{\Gamma_1\Gamma_2}\left(\frac{\sqrt{E_0E_F}}{\hbar\omega}\right)^5\frac{2}{\pi}g_3(\nu)C_\sigma, \tag{3.25}$$

$$g_3(\nu)=\frac{1}{\nu}\int_{1-\nu}^1\chi^{5/2}(\chi+\nu)^{3/2}\frac{\pi}{Q}d\chi. \tag{3.26}$$

In addition to the combinations e^2N/m and $(ea/\hbar)^2I_0$ mentioned earlier, Eqs. (3.23) and (3.25) contain the parameter

$$\frac{2\sqrt{E_0E_F}}{\hbar\omega}=\frac{v_F}{a\omega}. \tag{3.27}$$

In classical terms this parameter is the ratio of the rate of electron collisions with the surface of the sphere of radius a to the field frequency, while in quantum terms it is the ratio of the characteristic distance between the closest levels, $2\sqrt{E_0E_F}$, near the Fermi surface to the photon energy. Thus, the parameter (3.27) characterizes the size, or quantum, effects.

If $\hbar\omega\ll E_F$ holds, the values of the functions $g_1(\nu)$ and $g_3(\nu)$ are close to unity. This condition is barely satisfied in the spectral region where we can still ignore the absorption of light by the lattice ($E_F=5\times10^4$ cm^{-1} and $\omega=10^4$ cm^{-1}). The factor Q in (3.24) and (3.26) differs somewhat from state to state, but even in fairly accurate estimates we can take $Q=\pi$; in contrast to (3.17) and (3.21), which incorporate large powers of Q, here small inaccuracies in Q lead to small errors.

To make things more convenient in the discussion that follows, we combine the results into a single expression and keep only the first corrections in the small parameter Γ_2/ω:

$$D_\sigma = -\frac{e^2 N}{m\omega^2}\left\{\left[F_1 - i\left(\frac{2\Gamma_2}{\omega}F_1 + \frac{2\sqrt{E_0 E_F}}{\hbar\omega}g_1\right)\right]\right.$$
$$\times E_\sigma - \left(\frac{ea}{\hbar\omega}\right)^2 I_0\,\frac{\Gamma_2}{\Gamma_1}\left[F_3 - i\left(\frac{2\Gamma_2}{\omega}F_3\right.\right.$$
$$\left.\left.\left. + \frac{\omega^2}{(2\Gamma_2)^2}\left(\frac{2\sqrt{E_0 E_F}}{\hbar\omega}\right)^5 g_3\right)\right]C_\sigma\right\}. \tag{3.28}$$

Here the terms containing F_1, F_3 and g_1, g_3 are caused by the Drude and size effects, respectively. The real parts contribute to refraction and the imaginary parts, to absorption.

4. INDUCED VARIATION OF POPULATION

The problem of relaxation of a nonequilibrium electron distribution has a long history (see, e.g., Refs. 23–25), and lately has attracted the attention of researchers in connection with the experimental possibilities introduced by the technique of ultrashort pulses of laser light.[26,27] We believe that the problem of the hierarchy of processes of exchange of the energy taken by electrons from the field, such as thermalization of ''hot electrons,'' energy transport from electron to lattice, and anisotropy relaxation, has yet to be solved. For small particles the exchange of energy with the surrounding medium, or heat conductance, may also play a certain role. Apparently, the answers to all these questions depend, at least partially, on the type of induced distribution of electrons over the states. In view of what has been said in connection with the model (2.2), we analyze the distribution the electrons acquire as a result of their interaction with radiation.

In the first approximation in E, Eqs. (2.2) yield

$$\rho_{ij}^{(0)} = \delta_{ij}N_i + L_{ij}^{(0)}\sum_l G_{il}G_{lj}$$
$$\times[(L_{lj}^{(-1)} + L_{lj}^{(1)})N_{jl} - (L_{il}^{(-1)} + L_{il}^{(1)})N_{li}]. \tag{4.1}$$

Here the indices i, j, and l indicate sets of quantum numbers

$$i = n_i l_i M_i, \quad j = n_i l_i M_i', \quad l = n_l l_l M_l,$$

i.e., $\rho_{ij}^{(0)}$ describes the stationary population and Zeeman coherence of a state with energy E_i. As before, for the levels $E_i > E_F$ and $E_l < E_F$ we assume that $N_i = 0$ and $N_l = 1$ and that N_α is independent of M_α. Using the expression (2.11) for $L_{\alpha\beta}^{(s)}$, we arrive at the following formula:

$$\rho_{ij}^{(0)} = \frac{2\Gamma_2}{\Gamma_1}\sum_l G_{il}G_{lj}\left[\frac{1}{(\omega_{il}-\omega)^2+\Gamma_2^2}\right.$$
$$\left. + \frac{1}{(\omega_{il}+\omega)^2+\Gamma_2^2}\right]. \tag{4.2}$$

A similar expression exists for levels below the Fermi surface. In this case we have $E_i = E_j < E_F$, $N_i = 1$, $E_l > E_F$, $N_l = 1$, and

$$\rho_{ij}^{(0)} = \delta_{M_i M_i'} - \frac{2\Gamma_2}{\Gamma_1}\sum_l G_{il}G_{lj}$$
$$\times\left[\frac{1}{(\omega_{li}-\omega)^2+\Gamma_2^2} + \frac{1}{(\omega_{li}+\omega)^2+\Gamma_2^2}\right]. \tag{4.3}$$

Summation over M_l in Eqs. (4.2) and (4.3) involves only the vector addition coefficients in $G_{il}G_{lj}$, and the result is

$$\sum_{M_l} G_{il}G_{lj} = \frac{1}{3}\left|\frac{d_{il}}{2\hbar}\right|^2 I_0 \sum_{\kappa q}(-1)^{l_i - M_i'}\langle l_i M_i l_i$$
$$- M_i'|\kappa q\rangle I(\kappa q)a_{il\kappa}, \tag{4.4}$$
$$a_{il\kappa} = \sqrt{3}(-1)^{l_i + l_l + \kappa + 1}\begin{Bmatrix} 1 & 1 & \kappa \\ l_i & l_i & l_l \end{Bmatrix},$$

with the polarization tensor $I(\kappa q)$ defined in (2.16). Convenient characteristics of states are the polarization moments $\rho_i(\kappa q)$ of states i, linked to $\rho_i(M_i M_i')$ by the following relationships:[13–17]

$$\rho_i(\kappa q) = \sum_{M_i M_i'}(-1)^{l_i - M_i'}\langle l_i M_i l_i - M_i'|\kappa q\rangle\rho_i(M_i M_i'),$$
$$\rho_i(M_i M_i') = \sum_{\kappa q}(-1)^{l_i - M_i'}\langle l_i M_i l_i - M_i'|\kappa q\rangle\rho_i(\kappa q). \tag{4.5}$$

Combining (4.2) and (4.4), we arrive at the following formula for $\rho_i(\kappa q)$:

$$\rho_i(\kappa q) = I_0\frac{2\Gamma_2}{3\Gamma_1}\sum_{n_l l_l}\left|\frac{d_{il}}{2\hbar}\right|^2\left[\frac{1}{(\omega_{il}-\omega)^2+\Gamma_2^2}\right.$$
$$\left. + \frac{1}{(\omega_{il}+\omega)^2+\Gamma_2^2}\right]I(\kappa q)a_{il\kappa}. \tag{4.6}$$

For $l_l \gg 1$ the coefficients $a_{il\kappa}$ are

$$a_{il\kappa} = \frac{a_\kappa}{\sqrt{2l_i}}, \quad a_0 = 1, \quad a_1 = \pm\frac{1}{\sqrt{2}}, \quad a_2 = \frac{1}{\sqrt{10}}, \tag{4.7}$$

where the different signs of a_1 correspond to $l_l = l_i \mp 1$, respectively. Thus, the factor $a_{il\kappa}I(\kappa q)$ is practically independent of l_l and determines the polarization features of the moments $\rho_i(\kappa q)$. For linear polarization we have

$$a_0 I(00) = 1, \quad a_2 I(20) = -\frac{1}{\sqrt{5}}, \tag{4.8}$$

and for circular polarization,

$$a_0 I(00) = 1, \quad a_1 I(10) = \pm\frac{\sqrt{3}}{2}, \quad a_2 I(20) = \frac{1}{2\sqrt{5}}. \tag{4.9}$$

Note that the numerical values (4.8) and (4.9) are related to the relaxation model (2.2), within which the relaxation rates Γ_1 are the same for all $\rho_i(\kappa q)$. In more complicated (and realistic) models, the Γ_1 depend on κ, and the polarization moments of different ranks are related differently (see, e.g., Sec. 10 in Ref. 13 and Sec. 4.1 in Ref. 14).

Now we insert the explicit expression (2.6) for d_{il}, with the result that

$$\rho_i(\kappa q)=\left(\frac{ea}{\hbar}\right)^2 I_0\frac{8\Gamma_2}{3\Gamma_1}\frac{a_\kappa I(\kappa q)}{\sqrt{2l_i}}\sum_{n_l l_l}\frac{\alpha_i^2\alpha_l^2 l_{\max}^{il}}{(\alpha_i^2-\alpha_l^2)^4}\delta_{l_l l_i\pm 1}$$
$$\times\left\{\frac{1}{(\omega_{il}-\omega)^2+\Gamma_2^2}+\frac{1}{(\omega_{il}+\omega)^2+\Gamma_2^2}\right\}. \qquad (4.10)$$

We sum over $n_l l_l$, assuming that the regions of low frequencies and resonance frequencies do not overlap (as in Sec. 3). In the low-frequency region we keep only one term in the series, the one with the minimum value of $\alpha_i-\alpha_l$:

$$\alpha_i-\alpha_l=P, \quad l_l=l_i-1; \quad \alpha_i-\alpha_l=Q-P, \quad l_l=l_i+1.$$

The other terms of the series can be taken into account via formula (B5), but qualitatively the result is the same as without these terms, so that we will not write the corresponding corrections. We arrive at the following relationship:

$$\rho_i(\kappa q)=\left(\frac{ea}{\hbar\omega}\right)^2 I_0\frac{\Gamma_2}{3\Gamma_1}\sqrt{\frac{l_i}{2}}\,|a_\kappa|I(\kappa q)\left\{\left[\frac{E_i-E_F}{2\sqrt{E_0E_F}}+P\right]^{-4}\right.$$
$$+\left[\frac{E_i-E_F}{2\sqrt{E_0E_F}}+Q-P\right]^{-4}+[1+(-1)^\kappa]$$
$$\left.\times\frac{\pi}{2Q}\frac{\omega}{\Gamma_2}\left(\frac{\sqrt{E_0E_F}}{\hbar\omega}\right)^3\frac{E_i\sqrt{E_i-\hbar\omega}}{E_F^{3/2}}\right\}. \qquad (4.11)$$

The saturation parameter is still $(ea/\hbar\omega)^2 I_0$. What is most important in (4.11) is the dependence on the energy E_i. Inside the braces there are terms of different types. The first two terms rapidly decrease (as fourth-order hyperbolas) when the state energies E_i exceed the Fermi energy. The hyperbolic term with P corresponds to the transition $l_l\to l_i=l_l+1$ and the term with $Q-P$ corresponds to the transition $l_l\to l_l=l_l-1$. The hyperbolic terms, which can be called Drude terms, practically disappear in the energy layer $E_i-E_F=2\sqrt{E_0E_F}$. The third term in (4.11) corresponds to size effects. Its dependence on E_i is much smoother, since

$$E_i-E_F\lesssim\hbar\omega\ll E_F. \qquad (4.12)$$

The relationship between the Drude and size parts of the induced variation of $\rho_i(\kappa q)$ (on the Fermi surface) is specified by the parameter

$$K_2=\frac{\pi P^4}{Q}\frac{\omega}{\Gamma_2}\left(\frac{E_0E_F}{\hbar\omega}\right)^3\propto\frac{E_F^{3/2}}{\Gamma_2}\frac{\lambda^2}{a^3}, \qquad (4.13)$$

which is usually smaller than unity (if $\omega=10^4$ cm^{-1}, $\Gamma_2=10^2$ cm^{-1}, and $v_F=1.4\times10^8$ cm s^{-1}, then $K_2=0.03$). But at $E_i-E_F=\hbar\omega$, at the limit of the induced variation of $\rho_i(\kappa q)$, the size part is considerably larger than the Drude part.

In states with $E_i<E_F$, the dependence on E_i for the induced variation of $\rho_i(\kappa q)$ is similar. A rough sketch of the $\rho_i(\kappa q)$ dependence on E_i is given in Fig. 4.

Qualitatively, the Drude part of the field variation of $\rho_i(\kappa q)$ as a function of E_i is similar to the temperature deformation of the Fermi distribution. Bearing in mind that the thickness of the hyperbolic layer ($2\sqrt{E_0E_F}\sim10^2-10^3$ cm^{-1})

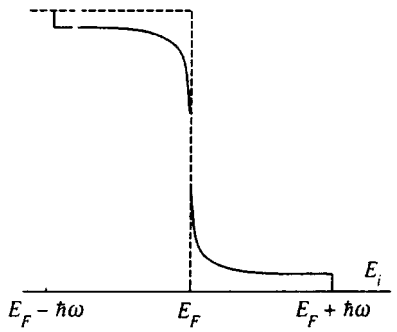

FIG. 4. The energy distribution of electrons with allowance for stimulated transitions. The dashed curve represents the Fermi distribution at absolute zero.

can exceed the thermal energy by a factor of ten at the most, there are grounds to believe that this part of $\rho_i(\kappa q)$ is directly transformed as a result of relaxation into the quasi-equilibrium distribution with a certain effective temperature. On the other hand, the size part has a width of order $\hbar\omega\gg T$ on the energy scale. Hence its relaxation, the relaxation of ''hot electrons,'' may proceed in a more complicated, multistage, way.

5. DISCUSSION

Let us discuss the consequences of choosing the interaction Hamiltonian in the form (1.2). The matrix elements of the operators (1.1) and (1.2) differ by the factor ω_{mn}/ω, so that if we were to use the operator $\mathbf{p}\cdot\mathbf{A}$ instead of $\mathbf{d}\cdot\mathbf{E}$, the terms in the series (3.3) and (3.5) would acquire the additional factors

$$\left(\frac{\omega_{mn}}{\omega}\right)^2=\left[\frac{E_0(\alpha_m^2-\alpha_n^2)}{\hbar\omega}\right]^2,$$
$$\left(\frac{\omega_{mn}}{\omega}\right)^4=\left[\frac{E_0(\alpha_m^2-\alpha_n^2)}{\hbar\omega}\right]^4, \qquad (5.1)$$

respectively. In resonance, $\omega_{mn}=\omega$, the factors (5.1) contribute nothing new. However, in the low-frequency region ($\omega_{mn}\ll\omega$) the situation is different: the combination $\alpha_m^2\alpha_n^2/(\alpha_m^2-\alpha_n^2)$ remains in the terms of the series (3.3), and without a resonant denominator the series does not converge. Hence, if we wish to use the interaction Hamiltonian in the form (1.1), summation over the region of low and resonance Bohr frequencies must be done simultaneously. Hache *et al.*[6] bypass the problem by introducing *a priori* a Drude term identical to the one in the case of a massive specimen. Such an approach is inconsistent, and within it one cannot expect to correctly calculate the nonlinear components of the polarization of a ball. In contrast to (1.1), the operator $\mathbf{d}\cdot\mathbf{E}$ produces a rapidly converging series (3.3). Hence, for our problem the most suitable is the interaction Hamiltonian in the form (1.2). The preferability of our approach has been established in nonlinear optical phenomena as well.[8,10]

The apparatus of irreducible spherical operators, which became widespread in atomic spectroscopy after the publication of Sobelman's classical monograph,[12] proved its effectiveness in our problem, too. In particular, summation over

the magnetic quantum numbers can easily be done in the linear, cubic, and, generally, any approximation in the field amplitude. The final expressions are obtained in a more compact form than if we use the Cartesian system of coordinates, as was done in previous papers (see Ref. 6 and the literature cited therein). The vector C_σ (3.20) and the tensor $I(\kappa q)$ are formed by the standard operation of the tensor multiplication of operators, (see, e.g., Sec. 14 in Ref. 12). The most frequently used are the linear and circular polarizations of radiation. For linear polarization (the quantization axis directed along **E**) calculations yield

$$C_\sigma = C_0 \delta_{\sigma 0} = \frac{2}{5} E \delta_{\sigma 0}, \tag{5.2}$$

and for circular polarization (the z axisdirected along the wave vector) we have

$$C_\sigma = C_1 \delta_{\sigma 1} = \frac{3}{5} E \delta_{\sigma 1}. \tag{5.3}$$

The ratios of the isotropic ($\kappa = 0$) to the anisotropic ($k = 1$ and 2) parts in these cases are 5:1 and 5:4, respectively.

In the Drude and size terms of D_σ, the prevailing parts are the real and imaginary, respectively, which is due to the general fact according to which forced vibration in resonant conditions are characterized by a $\pi/2$ shift in phase in relation to the driving force, while in nonresonant conditions this phase shift is 0 or π. This remark refers to both linear and nonlinear components of D_σ (the factor i in Eqs. (3.23) and (3.25)).

According to (3.28), there are two small parameters that determine the relationship between the Drude and size parts:

$$\frac{\Gamma_2}{\omega}, \quad \frac{2\sqrt{E_0 E_F}}{\hbar \omega} = \frac{v_F}{a\omega}. \tag{5.4}$$

The ratio of these parameters,

$$K_1 = \frac{v_F}{a\Gamma_2}, \tag{5.5}$$

plays an important role in the imaginary part of $D_\sigma^{(1)}$: it is frequency-independent, is determined only by the parameters of the metal, and at $a = 100$ Å is usually much larger than unity (e.g., for silver at $v_F = 1.4 \times 10^8$ cm s^{-1}, $\Gamma_2 = 10^2$ cm^{-1}, and $a = 100$ Å we have $K_1 = 8$).

The same parameters enter into the imaginary part of $D_\sigma^{(3)}$ as a different ratio:

$$K_3 = \frac{1}{8}\left(\frac{\omega}{\Gamma_2}\right)^3 \left(\frac{2\sqrt{E_0 E_F}}{\hbar \omega}\right)^5 = \frac{v_F^5}{8\Gamma_2^3 a^5 \omega^2}. \tag{5.6}$$

Here the parameters, including the light frequency, are raised to higher powers, so that K_3 can be either larger or smaller than unity. For instance, for $a = 100$ Å, $\Gamma_2 = 10^2$ cm^{-1}, and $\lambda = 1$ μm we have $K_3 = 0.3$, but variations in a and ω by a factor of 1.5–2 change the value of K_3 by a factor of ten. Thus, depending on the conditions, either the Drude or the size (quantum) effect plays an important role.

In the problem of induced variation of population (Sec. 4) we encountered the parameter K_2 (4.13), which for the above values of the numerical parameters is much smaller than unity, i.e., near the Fermi surface the main nonlinear effect is the Drude effect. However, at $E_i - E_F \approx \hbar\omega$, the size effect dominates (see Fig. 4).

In our opinion, an important result is the fact that the Drude nonlinearity depends on the frequency of the field. According to (3.19), the real part of $D_{\sigma D}^{(3)}$ is inversely proportional to the fourth power of the frequency. Physically this is quite obvious, since the factor $(\omega + i\Gamma_2)^{-2}$ specifies the frequency dependence of the ''absorption profile'' of the free electron (as it does in $D_{\sigma D}^{(1)}$), while the factor $(ea/\hbar\omega)^2 I_0$ is the saturation parameter, which determines the variation in level population. In other words, there is direct similarity with the Karplus–Schwinger formula,[28] which describes the saturation effect in the two-level model. Methodologically, the $D_{\sigma D}^{(3)} \propto \omega^{-4}$ dependence is a direct consequence of choosing the interaction operator in the form $\mathbf{d}\cdot\mathbf{E}$. In the theory of Hache *et al.*,[6] based on the Hamiltonian (1.1), the dependence is quite different: $D_{\sigma D}^{(3)} \propto \omega^{-7}$. This, in particular, means that the conclusions drawn in Ref. 6 about the relative role of the Drude and size components do not correspond to the true situation. For the same reason we doubt that the comparison of the theoretical and experimental results done in Ref. 6 is meaningful.

The analog of the saturation parameter for the resonant contribution in $D_\sigma^{(3)}$ is the combination $(ea/\hbar)^2 I_0/\Gamma_1\Gamma_2$, which instead of ω^2 contains the population and polarization relaxation rates Γ_1 and Γ_2. This structure of the parameter is in full agreement with the idea of saturation in resonant absorption.

The frequency dependence of the size component, $D_{\sigma a}^{(3)} \propto \omega^{-7}$, which coincides with the one obtained in Ref. 6, is caused by the decrease in the matrix elements d_{mn} with increasing natural frequencies ω_{mn}, which provide the resonant contribution to $D_{\sigma a}^{(3)}$ (see formula (2.6)).

Let us now estimate the numerical value of the saturation parameter $(ea/\hbar\omega)^2 I_0$, which characterizes the extent to which the nonlinear effects manifest themselves. Expressing $W = (c/8\pi) I_0$, a, and λ, respectively, in MW cm^{-2}, 10 nm, and μm,

$$I_0 = \frac{8\pi}{c} W = \frac{8\pi}{c}\, 10^{13} \tilde{W} \ \text{MW}\cdot\text{cm}^{-2},$$

$$a = 10^{-6}\tilde{a} \times 10 \ \text{nm}, \quad \lambda = 10^{-4}\tilde{\lambda} \ \mu\text{m},$$

we find that

$$\left(\frac{ea}{\hbar\omega}\right)^2 I_0 = 0.49 \times 10^{-3} \tilde{W}\tilde{a}^{\,2}\tilde{\lambda}^2. \tag{5.7}$$

Thus, for a radiation power $W = 1$ MW cm^{-2}, a particle radius $a = 10$ nm, and a wavelength $\lambda = 1$ μm the saturation parameter amounts to 0.05%.

There are two factors that enhance the nonlinear effects. According to (3.28), $D_\sigma^{(3)}$ is proportional to Γ_2/Γ_1, which can be of order 10 (Ref. 6). The second factor is related to the fact that the external field E_0 specifying W differs from the field E inside the ball, with the latter entering into the formulas for D_σ (see Refs. 6 and 29):

$$E=fE_0, \quad f=\left[1+\frac{\varepsilon-\varepsilon_0}{3\varepsilon_0}\right]^{-1}, \tag{5.8}$$

where ε and ε_0 are the dielectric constants of the ball and the surrounding medium. In the spectral region near the point where $\mathrm{Re}[1+(\varepsilon-\varepsilon_0)/3\varepsilon_0]=0$, the value of $|f|$ may be large. Since $f|f|^2$ enters into the cubic terms, the corresponding enhancement of the nonlinear effects may be considerable.

The overall result of our analysis is that the approach, ideas, and methods of nonlinear spectroscopy developed earlier for atoms and molecules[13,14] can be successfully applied to the degenerate electron gas in metallic particles. Consequently, it would be interesting to apply in this case the powerful method of a probe field, in which nonlinear interference effects lead to vivid polarization and spectroscopic phenomena.[3)]

The author is grateful to V. P. Safonov for discussing various aspects of the problem and for fruitful remarks. The research was sponsored by the Russian Fund for Fundamental Research (Grant No. 96-02-19331) and the International Science Foundation (Grant No. 86_p).

APPENDIX A

For the roots of Bessel functions we have the following formulas:[21]

$$\alpha_{nl}=\pi\left(n+\frac{l}{2}\right)-\frac{l(l+1)}{2\pi(n+l/2)} \times\left\{1+\frac{7l(l+1)-6}{3[2\pi(n+l/2)]^2}+\cdots\right\}, \quad n\gg l, \tag{A1}$$

$$\alpha_{nl}=l+\frac{1}{2}+Q_{1n}\left(l+\frac{1}{2}\right)^{1/3}+Q_{2n}\left(l+\frac{1}{2}\right)^{-1/3}+\cdots, \quad n\ll l. \tag{A2}$$

At $l=0$, Eq. (A1) is exact:

$$\alpha_{n0}=\pi n, \quad n=1, 2, 3,\ldots . \tag{A3}$$

When n and l vary within a limited interval, the following interpolation formula is true:

$$\alpha_{nl}=Q_{nl}n+P_{nl}l+R_{nl}, \tag{A4}$$

with the coefficients Q_{nl}, P_{nl}, and R_{nl} being weakly dependent on nl. The following inequalities also hold:

$$Q_{nl}\geqslant\pi, \quad 1<P_{nl}\leqslant\frac{\pi}{2}. \tag{A5}$$

Here $P_{nl}=\pi/2$ and $Q_{nl}=\pi$ hold only at $l=0$, and $P_{nl}=1$ holds when $l\to\infty$.

For $\alpha_F=120$ on the Fermi surface in the interval $70\leqslant l\leqslant 110$, where thanks to the statistical weight $2l+1$ approximately half the states are present, we have

$$10>n>1, \quad 4<Q_{nl}<7, \quad 1.2>P_{nl}>1.03.$$

APPENDIX B

In the definition of the generalized Riemann zeta function,[19,20]

$$\zeta(s,v)=\sum_{k=0}^{\infty}(k+v)^{-s}, \tag{B1}$$

we isolate the first terms in the series, $k=0, 1,\ldots,m$, and in the other replace $k+v$ via the equality

$$\Gamma(s)(k+v)^{-s}=\int_0^{\infty}\exp[-(k+v)z]z^{s-1}dz \tag{B2}$$

and sum the series. Simple transformations lead to the following formula:

$$\zeta(s,v)=\sum_{k=0}^{m}(k+v)^{-s}+\frac{1}{(s-1)(m+v+1/2)} \times\left\{1-\frac{1}{\Gamma(s-1)}\int_0^{\infty}\left[1-\frac{u}{\sinh u}\right]e^{-\chi}\chi^{s-2}d\chi\right\}, \tag{B3}$$

where $u=\chi/(1+2m+2v)$. The coefficient of the expression in the braces coincides with the estimate of the remainder series by the integral (upper bound). The integral in the braces is positive, decreases with increasing m and v, and can be discarded:

$$\zeta(s,v)=\sum_{k=0}^{m}(k+v)^{-s}+(s-1)^{-1}\left(m+v+\frac{1}{2}\right)^{1-s}. \tag{B4}$$

For v small and s large, good accuracy is achieved even if we keep only one term (with $m=0$):

$$\zeta(s,v)=\frac{1}{v^s}\left\{1+\frac{v}{s-1}\left(\frac{v}{v+1/2}\right)^{s-1}\right\}. \tag{B5}$$

In the region $0<v<1$ and $s>2$, at $m=0$ and $m=1$ formula (B4) yields an error less than 2% and 0.3%, respectively.

[1)] With allowance for the electron spin, we could have normalized the N_j in (2.14) to 2. However, it has proved more convenient to introduce the statistical weight and allow for spin later, in Eq. (3.14).

[2)] Note that Gor'kov and Éliashberg[5] examined the ultraquantum case $\hbar\omega<2\sqrt{E_0E_F}$, in which these two regions are not separated.

[3)] In calculating the influence of the probe field, Bigot *et al.*[27] ignored nonlinear interference effects.

[1] J. C. Maxwell-Garnett, Philos. Trans. R. Soc. London **203**, 385 (1904); **205**, 237 (1906).
[2] G. Mie, Ann. Phys. (Leipzig) **25**, 377 (1908).
[3] V. L. Ginzburg and G. P. Motulevich, Usp. Fiz. Nauk **55**, 469 (1955).
[4] A. Kabawata and R. Kubo, J. Phys. Soc. Jpn. **21**, 1765 (1966).
[5] L. P. Gor'kov and G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **48**, 1407 (1965) [Sov. Phys. JETP **21**, 940 (1965)].
[6] F. Hache, D. Ricard, and C. Flytzanis, J. Opt. Soc. Am. B **3**, 1647 (1986).
[7] N. Bloembergen, *Nonlinear Optics*, W. A. Benjamin, New York (1965).
[8] N. B. Delone and V. P. Kraĭnov, *Atoms in Strong Light Fields*, Springer, Berlin (1985).
[9] Yu. A. Il'inskiĭ and L. V. Keldysh, *Electromagnetic Response of Material Media*, Plenum, New York (1994).
[10] F. V. Bunkin, Zh. Éksp. Teor. Fiz. **50**, 1685 (1966) [Sov. Phys. JETP **23**, 1121 (1966)].
[11] V. M. Galitskiĭ, B. M. Karnakov, and V. I. Kogan, *Problems in Quantum Mechanics* [in Russian], Nauka, Moscow (1981).
[12] I. I. Sobelman, *Introduction to the Theory of Atomic Spectra*, Pergamon Press, Oxford (1973); *Atomic Spectra and Radiative Transitions*, Springer, Berlin (1979).
[13] S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances*

in the Spectra of Atoms and Molecules [in Russian], Nauka, Novosibirsk (1979).
[14] S. G. Rautian and A. M. Shalagin, *Kinetic Problems of Non-linear Spectroscopy*, North-Holland, Amsterdam (1991).
[15] A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, Princeton Univ. Press, Princeton, N.J. (1957).
[16] M. I. D'yakonov and V. I. Perel', Zh. Éksp. Teor. Fiz. **47**, 1483 (1964) [Sov. Phys. JETP **20**, 997 (1965)].
[17] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum*, World Scientific, Singapore (1987).
[18] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[19] A. Erdélyi, *Higher Transcendental Functions*, (Bateman Project) Vol. 1, McGraw-Hill, New York (1953).
[20] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th edn., Cambridge Univ. Press, Cambridge (1944).
[21] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, National Bureau of Standards Applied Mathematics Series 55, Washington, D.C. (1964).
[22] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Part 1, 3 rd ed., Pergamon Press, Oxford (1980).
[23] A. H. Wilson, *The Theory of Metals*, 2nd edn., Cambridge Univ. Press, Cambridge (1954).
[24] M. I. Kaganov, I. M. Lifshits, and L. V. Tanatarov, Zh. Éksp. Teor. Fiz. **31**, 232 (1956) [Sov. Phys. JETP **4**, 173 (1957)].
[25] E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics*, Pergamon Press, Oxford (1981).
[26] R. N. M. Groeneveld, R. Sprik, and A. Lagendijk, Phys. Rev. B **51**, 11 433 (1995).
[27] J.-Y. Bigot, J.-C. Merle, O. Cregut, and A. Daunois, Phys. Rev. Lett. **75**, 4702 (1995).
[28] R. Karplus and J. Schwinger, Phys. Rev. **73**, 1020 (1948).
[29] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, Oxford (1984).

Translated by Eugene Yankovsky

Anomalous drift of resonant particles in a buffer medium under pressure of light

L. V. Il'ichev and A. I. Parkhomenko

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

(Submitted 31 December 1996)
Zh. Éksp. Teor. Fiz. **112**, 856–868 (September 1997)

We study theoretically the drift of resonant particles in a buffer medium when a traveling light wave impinges on the medium, with allowance for the velocity dependence of the transport collision rate. When the pressure of light dominates over the light-induced drift (low pressure of the buffer gas or the drift of conduction electrons in semiconductors), we discover a new sudden transformation of the spectral dependence of the drift velocity of the resonant particles: Instead of the ordinary bell-shaped function representing the velocity spectrum we have a double-humped curve with deep dip at the center of the absorption line. We show that the largest transformation of the drift velocity spectrum occurs in the atmosphere of a heavy buffer gas in the case of Coulomb interaction between the resonant and buffer particles. The transformation effect is caused by the variation of the transport rate of the collisions of the resonant and buffer particles due to the recoil effect in the absorption of radiation. © *1997 American Institute of Physics.* [S1063-7761(97)00709-9]

1. INTRODUCTION

The two effects that have attracted the most attention in the physics of selective action of radiation on the translational motion of resonant particles in a buffer medium (atoms and molecules in a gaseous medium, ions in a plasma, and conduction electrons in a solid, with phonons, impurities, and lattice defects acting as the buffer gas) are the resonant light pressure[1–3] and the effect of light-induced drift[1)] discovered in 1979 (Refs. 4–11).

Qualitatively, light pressure and light-induced drift manifest themselves in the same way, i.e., fluxes of absorbing particles are generated in gases and solids. Depending on the object being investigated and the prevailing conditions, the researcher is forced to consider the combined action of these effects (in the case of their ''competition'') or may ignore one effect.

Light pressure may dominate over light-induced drift if the light acts on the conduction electrons in semiconductors (photocurrent generation) or if the pressure of the buffer gas in the gaseous medium is low.

An important characteristic of the manifestation of light pressure is the spectral dependence of the drift velocity **u** of the absorbing particles in a buffer medium that is under the pressure of light (e.g., the spectral dependence of the electron drift velocity in semiconductors is directly related to the spectral characteristic of the sensitivity of a semiconductor light detector based on photon drag of electrons). When the transport rate of collisions of the resonant and buffer particles is independent of the magnitude of the velocity $v=|\mathbf{v}|$, i.e., $\nu(v)=\nu=\text{const}$, the drift velocity **u** is given by the standard relation (see, e.g., Ref. 7):

$$\mathbf{u}=\frac{\mathbf{k}}{k}v_r\frac{P}{\nu}, \tag{1}$$

where P is the probability for the absorption of radiation by a particle per unit time (absorption rate), **k** is the wave vector of the radiation, and $v_r=\hbar k/M$ is the velocity of recoil in the absorption (emission) of a photon, here M is the mass of the absorbing particles. Thus, with $\nu(v)=\text{const}$, the spectrum of the drift velocity **u** fully matches the spectral characteristic of the radiation absorption. When two-level particles are involved (which is the case studied in the present paper), the absorption rate P (the velocity **u**) is represented by the well-known bell-shaped curve known as the Voigt profile.

We know of no detailed analysis of light pressure when $\nu(v)\neq\text{const}$. It might seem that the natural generalization

$$\mathbf{u}=\frac{\mathbf{k}}{k}v_r\int\frac{P(\mathbf{v})}{\nu(v)}d\mathbf{v} \tag{2}$$

of Eq. (1) for the drift velocity in the case where $\nu(v)\neq\text{const}$ is sufficient and that no new features can emerge [here $P(\mathbf{v})$ is the probability for the radiation absorption per unit time by particles with a fixed velocity **v**, so that $P\equiv\int P(\mathbf{v})d\mathbf{v}$]. The spectrum of the velocity (2) is also represented by a bell-shaped curve, narrowed or expanded in comparison to the case where $\nu(v)=\text{const}$, depending on whether the collision rate $\nu(v)$ increases or decreases with increasing v.

Allowance for variations in the transport collision rate $\nu(v)$ caused by the particle acquiring an additional velocity v_r in the direction of the wave vector **k** would seem to lack all meaning because of the small [$\sim v_r/\overline{v}\ll 1$, where $\overline{v}$ is the mean thermal velocity of the particles, which determines the characteristic scale of variation of $\nu(v)$] correction to the collision rate. Nevertheless, it was found that this small correction leads to a drastic qualitative change in the spectrum of the drift velocity **u**: instead of the usual bell-shaped curve, the spectrum profile is a double-humped curve with a deep dip at the center of the absorption line.

In this paper we study theoretically this unexpectedly sudden transformation of the drift velocity spectrum.

2. THE QUALITATIVE PICTURE

What is the physics of the appearance of an anomalous spectrum in the drift velocity? Let us examine the behavior of two-level particles in an atmosphere of a buffer gas. To simplify matters, we assume that the transport rates $\nu_m(v)$ and $\nu_n(v)$ of collisions of the resonant particles on the combining levels and the buffer medium are equal, $\nu_m(v)=\nu(v)=\nu(v)$, and that the particles lose all memory about their velocity in the course of a single collision. We also assume that during its lifetime in the excited state m a particle experiences many collisions. Let us examine the case of Doppler absorption-line broadening, where the only particles that interact with the radiation are those whose velocity projection v_z on the wave vector $\mathbf{k}$ is close to $v_0=\Omega/k$, with $\Omega=\omega-\omega_{mn}$ the detuning of the radiation frequency ω from the center of the absorption line, ω_{mn}. The other two velocity projections, v_x and v_y, which are orthogonal to $\mathbf{k}$, can be arbitrary. Since the distribution over the velocities v_x and v_y is practically an equilibrium one,[7] we average the problem over v_x and v_y before we make any estimates. In this way the three-dimensional problem is effectively reduced to a one-dimensional problem with a transport rate depending only on the absolute value of the velocity projection v_z.

Let us examine the ''fate'' of a single particle. We take a particle that initially had an arbitrary velocity projection v_z. As a result of collisions, the particle finds itself in the resonant velocity region with $v_z=v_0$. Under excitation (the probability for the absorption of radiation by the particle per unit time, or the absorption rate, is P) the particle acquires, due to the recoil effect, an additional velocity v_r in the direction coinciding with that of the wave vector $\mathbf{k}$, so that in the excited state m this particle has a velocity $v_z=v_0+v_r$. The transport paths $\mathbf{l}_n$ and $\mathbf{l}_m$ of the particle in the ground state and the excited state differ in view of the dependence of the transport collision rate on v_z:

$$\mathbf{l}_n=\frac{\mathbf{k}}{k}\frac{v_0}{\nu(v_0)}, \quad \mathbf{l}_m=\frac{\mathbf{k}}{k}\frac{v_0+v_r}{\nu(v_0+v_r)}. \tag{3}$$

Ultimately, the particle, which prior to excitation was traveling at random, acquires a directional displacement $\mathbf{l}=\mathbf{l}_m-\mathbf{l}_n$. This displacement process is repeated P times each second and leads to a drift of the resonant particles with an average velocity

$$\mathbf{u}=\mathbf{l}P. \tag{4}$$

Combining (3) and (4), we find

$$\mathbf{u}=\mathbf{u}_r+\mathbf{u}_a, \tag{5}$$

where

$$\mathbf{u}_r=\frac{\mathbf{k}}{k}v_r\frac{P}{\nu(v_0+v_r)}, \quad \mathbf{u}_a=-\frac{v_0}{v_r}\frac{\Delta\nu}{\nu}\mathbf{u}_r,$$

$$\frac{\Delta\nu}{\nu}\equiv\frac{\nu(v_0+v_r)-\nu(v_0)}{\nu(v_0)}. \tag{6}$$

In (5) the spectral dependence of the partial drift velocity $\mathbf{u}_r$ on the radiation frequency is a bell-shaped curve (just as the absorption rate P is) and is the well-known ''ordinary'' reaction of a system to spontaneous light pressure. The second term in (5), $\mathbf{u}_a$, is a reflection of the variation of the transport collision rate due to the recoil effect. It is responsible for the sudden transformation of the spectrum of the drift velocity $\mathbf{u}$.

Let us examine the behavior of the anomalous partial drift velocity $\mathbf{u}_a$. If the mass of the resonant part is smaller than, or comparable to, the mass of the buffer particles, the characteristic scale of variation of the transport rate $\nu(v_z)\equiv\nu(|v_z|)$ is of the order of the thermal velocity $\overline{v}$ (see Ref. 7), and for the factor $\Delta\nu/\nu$ we have the following estimate: $|\Delta\nu|/\nu\sim v_r/\overline{v}$. Thus, the anomalous partial drift velocity $\mathbf{u}_a$ may reach the value

$$|\mathbf{u}_a|\sim\frac{|\Omega|}{k\overline{v}}|\mathbf{u}_r|, \tag{7}$$

and for $|\Omega|>k\overline{v}$ its contribution to the total drift velocity $\mathbf{u}$ [Eq. (5)] can exceed the contribution $\mathbf{u}_r$ from the ''ordinary'' response of a medium to light pressure ($k\overline{v}$ is the Doppler width of the absorption line).

The direction of $\mathbf{u}_a$ specified by (6) depends on the sign of the product $v_0\Delta\nu\propto\Omega\Delta\nu$. If the transport rate $\nu(v_z)$ increases with $|v_z|$, then $\Delta\nu$ is positive for $v_0\propto\Omega>0$ and negative for $v_0\propto\Omega<0$. Thus, in this case $\mathbf{u}_a$ points in the opposite direction than $\mathbf{u}_r$ (in opposition to $\mathbf{k}$).

But if $\nu(v_z)$ decreases as $|v_z|$ grows, then $\Delta\nu$ is negative for $\Omega>0$ and positive for $\Omega<0$. In this case the directions of $\mathbf{u}_a$ and $\mathbf{u}_r$ coincide.

From (6) it follows that at the center of the absorption line ($\Omega=0$) $\mathbf{u}_a=0$. Thus, the spectral shape of $\mathbf{u}_a$ is a double-humped even function of the detuning Ω of the radiation frequency with zero at $\Omega=0$. When the directions of $\mathbf{u}_a$ and $\mathbf{u}_r$ coincide [i.e., when the collision rate $\nu(v_z)$ decreases with increasing $|v_z|$], the spectrum of the total drift velocity $\mathbf{u}$ [Eq. (5)] is a double-humped function with a dip at the center of the absorption line. But when $\mathbf{u}_a$ and $\mathbf{u}_r$ point in the opposite directions, the spectrum of the drift velocity $\mathbf{u}$ simply narrows without drastic changes in its shape.

3. GENERAL EXPRESSIONS

We study the interaction of a traveling monochromatic electromagnetic wave $\mathscr{E}=[E\exp(i\mathbf{k}\cdot\mathbf{r}-i\omega t)+\text{c.c.}]/2$ and two-level absorbing particles mixed with buffer particles. We ignore the collisions between the absorbing particles, assuming that the buffer gas concentration N_b is much higher than the concentration N of the absorbing gas. The following equations for the density matrix describe the particle–radiation interaction with allowance for the recoil effect:[7,12]

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\boldsymbol{\nabla}+\Gamma_m\right)\rho_m(\mathbf{v})=S_m(\mathbf{v})+NP(\mathbf{v}-\boldsymbol{\xi}),$$

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\boldsymbol{\nabla}\right)\rho_n(\mathbf{v})=S_n(\mathbf{v})+\hat{\Gamma}_m\rho_m(\mathbf{v})-NP(\mathbf{v}+\boldsymbol{\xi}),$$

$$\left[\frac{\partial}{\partial t}+\mathbf{v}\cdot\boldsymbol{\nabla}+\frac{\Gamma_m}{2}-i(\Omega_0-\mathbf{k}\cdot\mathbf{v})\right]\rho_{mn}(\mathbf{v})=S_{mn}(\mathbf{v})$$
$$+iG[\rho_n(\mathbf{v}-\boldsymbol{\xi})-\rho_m(\mathbf{v}+\boldsymbol{\xi})], \tag{8}$$

where

$$NP(\mathbf{v}) = -2\,\mathrm{Re}[iG^*\rho_{mn}(\mathbf{v})], \quad \boldsymbol{\xi} = \frac{\hbar\mathbf{k}}{2M},$$

$$\hat{\Gamma}_m\rho_m(\mathbf{v}) = \frac{\Gamma_m}{4\pi}\int \rho_m(\mathbf{v}+2\xi\mathbf{n})d\mathbf{n},$$

$$G = \frac{Ed_{mn}}{2\hbar}, \quad \Omega_0 = \omega - \omega_{mn}. \tag{9}$$

Here $\rho_n(\mathbf{v})$ and $\rho_m(\mathbf{v})$ are the velocity distributions of the particles at the ground level (n) and the excited level (m), $N = N_m + N_n$ is the concentration of the absorbing particles $(N_i = \int \rho_i(\mathbf{v})d\mathbf{v})$, $S_i(\mathbf{v})$ and $S_{mn}(\mathbf{v})$ are the collision integrals, d_{mn} is the matrix element of the dipole moment of the $m-n$ transition, ω_{mn} is the frequency of the $m-n$ transition, M is the mass of the absorbing particles, Γ_m is the rate of spontaneous relaxation of the excited level m, $\hat{\Gamma}_m\rho_m(\mathbf{v})$ is an integral operator that takes into account the radiative transport of particles from the excited level to the ground level with allowance for variations in particle velocity due to recoil in spontaneous emission, $\mathbf{n}$ is a unit vector specifying the direction of spontaneous emission, and $P(\mathbf{v})$ is the rate of absorption (absorption probability per unit time) of radiation by particles with a fixed velocity $\mathbf{v}$.

In the absence of phase memory in collisions, the off-diagonal collision integral has the form

$$S_{mn}(\mathbf{v}) = -[\nu_{mn}(v) + i\Delta_{mn}(v)]\rho_{mn}(\mathbf{v}), \tag{10}$$

where $\nu_{mn}(v)$ and $\Delta_{mn}(v)$ are the collision broadening and shift of the levels, respectively. Combining the last equation in (8) with (10), we find that under stationary and spatially homogeneous conditions the absorption rate $P(\mathbf{v})$ specified in (9) is

$$NP(\mathbf{v}) = 2|G|^2 Y(\mathbf{v})[\rho_n(\mathbf{v}-\boldsymbol{\xi}) - \rho_m(\mathbf{v}+\boldsymbol{\xi})], \tag{11}$$

where

$$Y(\mathbf{v}) = \frac{\Gamma(v)}{\Gamma^2(v) + [\Omega(v) - \mathbf{k}\cdot\mathbf{v}]^2},$$

$$\Gamma(v) = \frac{\Gamma_m}{2} + \nu_{mn}(v), \quad \Omega(v) = \Omega_0 - \Delta_{mn}(v). \tag{12}$$

For the diagonal collision integral we use the model of particle ''arrival'' that is isotropic in velocities:

$$S_i(\mathbf{v}) = -\nu_i(v)\rho_i(\mathbf{v}) + S_i^{(2)}(v), \tag{13}$$

where the ''arrival'' term $S_i^{(2)}(v)$ is a function of the absolute value of velocity, $v = |\mathbf{v}|$. The collisional model allows for a frequency dependence of the collision rate and at the same time makes it possible to solve the problem analytically. Note that $\nu_i(v)$ in (13) is the transport collision rate. Indeed, the force of internal friction, $\mathbf{F}_i$, which originates in the collisions of resonant particles in state i with the buffer gas, is given, on the one hand, by the standard expression in terms of the collision integral,

$$\mathbf{F}_i = M\int \mathbf{v}S_i(\mathbf{v})d\mathbf{v}. \tag{14}$$

On the other hand, the same force can be expressed as follows:[7]

$$\mathbf{F}_i = -M\int \nu_{i\ \mathrm{tr}}(\mathbf{v})\mathbf{v}\rho_i(\mathbf{v})d\mathbf{v}, \tag{15}$$

where $\nu_{i\ \mathrm{tr}}$ is the transport collision rate. The product $\mathbf{v}\rho_i(\mathbf{v})$ in the integrand in (15) is the flux density (in $\mathbf{v}$-space) of the particles with velocities $\mathbf{v}$, and $M\nu_{i\ \mathrm{tr}}(\mathbf{v})\mathbf{v}\rho_i(\mathbf{v})$ is the density of the friction force acting on the flux. Plugging (13) into (14) and comparing the result with (15), we conclude that $\nu_i(v)$ in (13) is the transport collision rate.

Let us find the velocity of drift of the resonant particles,

$$\mathbf{u} \equiv \frac{\mathbf{j}_m + \mathbf{j}_n}{N}, \quad \mathbf{j}_i = \int \mathbf{v}\rho_i(\mathbf{v})d\mathbf{v}, \tag{16}$$

where $\mathbf{j}_i$ is the flux of particles in state i. In stationary and spatially homogeneous conditions, by combining the first two equations in (8) with (13) we obtain the following expressions for the particle fluxes $\mathbf{j}_i$:

$$\mathbf{j}_m = N\int \mathbf{v}\frac{P(\mathbf{v}-\boldsymbol{\xi})}{\Gamma_m + \nu_m(v)}d\mathbf{v},$$

$$\mathbf{j}_n = \frac{\Gamma_m}{4\pi}\int \mathbf{v}\frac{\rho_m(\mathbf{v}+2\xi\mathbf{n})}{\nu_n(v)}d\mathbf{n}d\mathbf{v} - N\int \mathbf{v}\frac{P(\mathbf{v}+\boldsymbol{\xi})}{\nu_n(v)}d\mathbf{v}. \tag{17}$$

At $\nu_n(v) = \mathrm{const}$, the recoil effect does not influence the integral term in the expression (17) for $\mathbf{j}_n$. This follows from the fact that a spontaneous transition from level m to level n does not, on the average, change the particle momentum, so that

$$\frac{1}{4\pi}\int \mathbf{v}\rho_m(\mathbf{v}+2\xi\mathbf{n})d\mathbf{n}d\mathbf{v} = \int \mathbf{v}\rho_m(\mathbf{v})d\mathbf{v}. \tag{18}$$

In calculating the drift velocity we assume that, in general, $\nu_i(v) \neq \mathrm{const}$, and the recoil effect in the first integral term in the expression for $\mathbf{j}_n$ in (17) can be ignored, as analysis shows, only if

$$\frac{\xi}{\Delta v_n} \ll 1, \tag{19}$$

where Δv_n is the characteristic scale of variation of the transport rate $\nu_n(v)$.

If condition (19) is met, we obtain at the following expression for the drift velocity $\mathbf{u}$ defined in (16):

$$\mathbf{u} = \int \mathbf{v}\left[\frac{\nu_n(v) - \nu_m(v)}{\nu_n(v)}\frac{P(\mathbf{v}-\boldsymbol{\xi})}{\Gamma_m + \nu_m(v)} + \frac{P(\mathbf{v}-\boldsymbol{\xi}) - P(\mathbf{v}+\boldsymbol{\xi})}{\nu_n(v)}\right]d\mathbf{v}. \tag{20}$$

According to (20), the resonant particle flux generated by the radiation consists of two parts: the flux caused by the radiative force due to the recoil effect (the second term in the square brackets, which is nonzero only if we allow for the recoil effect), and the flux caused by the asymmetry of the function $P(\mathbf{v}-\boldsymbol{\xi})$ and the difference of the transport collision rates for levels m and n. If we ignore recoil $(\boldsymbol{\xi} = 0)$, Eq. (20) describes ordinary light-induced drift.

4. LOW RADIATION INTENSITIES

To calculate the drift velocity (20), we limit the discussion to the case of low radiation intensities:

$$\frac{2|G|^2}{\Gamma_m} \ll \Gamma + k\overline{v}, \quad \frac{\overline{v}\,\nu_n}{\xi\Gamma_m}(\Gamma + k\overline{v}), \quad \Gamma\left(1+\frac{\nu_m}{\Gamma_m}\right), \tag{21}$$

where ν_i and Γ are the characteristic (average) values of $\nu_i(v)$ and $\Gamma(v)$, and $\overline{v}=\sqrt{2k_BT/M}$ is the most probable velocity of the absorbing particles; here is T the temperature, and k_B the Boltzmann constant. If conditions (21) are met, then in (11) we can ignore the population of the excited level $[\rho_m(\mathbf{v})=0]$, and the velocity distribution of populations in the ground state can be assumed to be close to the Maxwellian distribution $[\rho_n(\mathbf{v})=NW(\mathbf{v})$, where $W(\mathbf{v})$ is the Maxwellian distribution]. In an approximation that is linear in the parameter $\xi/\overline{v}$, form Eq. (11) we find

$$P(\mathbf{v})=P_0(\mathbf{v})\left(1+\frac{2\mathbf{v}\cdot\boldsymbol{\xi}}{\overline{v}^2}\right), \quad P_0(\mathbf{v})=2|G|^2Y(\mathbf{v})W(\mathbf{v}). \tag{22}$$

Substituting (22) into (20) and assuming that a condition of type (19) is also met for the transport collision rate $\nu_m(v)$, we obtain the following expression for the drift velocity of the resonant particles:

$$\mathbf{u}\equiv\frac{\mathbf{k}}{k}u, \quad u=u_l+u_r+u_{lr}+u_a, \tag{23}$$

where

$$u_l=\int u_z\tau(v)P_0(\mathbf{v})d\mathbf{v}, \quad u_r=2\xi\int\frac{P_0(\mathbf{v})}{\nu_n(v)}d\mathbf{v},$$

$$u_{lr}=\xi\int\tau(v)P_0(\mathbf{v})\left[1+\frac{2v_z^2}{\overline{v}^2}\right]d\mathbf{v},$$

$$u_a=-\frac{\xi}{\overline{v}}\int\frac{v_z^2}{v}\frac{P_0(\mathbf{v})}{\nu_n(v)}[2\alpha_n(v)+a(v)]d\mathbf{v}. \tag{24}$$

Here we have introduced the notation

$$q(v)=\frac{\nu_m(v)}{\Gamma_m+\nu_m(v)}\left[\frac{\Gamma_m+\nu_n(v)}{\Gamma_m+\nu_m(v)}\alpha_m(v)-\alpha_n(v)\right],$$

$$\tau(v)=\frac{\nu_n(v)-\nu_m(v)}{\nu_n(v)[\Gamma_m+\nu_m(v)]},$$

$$\alpha_i(v)=\overline{v}\frac{d}{dv}\ln\nu_i(v), \quad v_z=\frac{\mathbf{k}\cdot\mathbf{v}}{k}. \tag{25}$$

Thus, according to (23) and (24), the drift velocity (and hence the force acting on the resonant particles) consists of four terms, each of its own physical nature. The first term u_l in (23), which represents the asymmetry of the absorption rate $P_0(\mathbf{v})$ and the difference in the transport collision rates on the combining levels m and n $[\tau(v)\neq0]$, is the ordinary light-induced drift velocity. The second term u_r, which represents the recoil effect $(\boldsymbol{\xi}\neq0)$, is the ''ordinary'' reaction of the system to spontaneous light pressure. The third term u_{lr} represents the recoil effect and the difference in transport collision rates. In a strong field and at high buffer-gas pressures $(|\nu_m-\nu_n|>\Gamma_m)$, the term u_{lr} is responsible for the ''negative'' light pressure, which was discussed in Ref. 13. Finally, the fourth term u_a represents the recoil effect and the velocity dependence of the transport collision rates $\nu_i(v)$. If the frequencies are velocity-independent, $\nu_i(v)=$const, then $\alpha_i(v)=0$, and u_a vanishes.

Integrating (24) over the directions of $\mathbf{v}$, we obtain the final expressions for the partial drift velocities u_l, u_r, u_{lr}, and u_a

$$u_l=\overline{v}\frac{2\kappa\Gamma_m}{\sqrt{\pi}}\int_0^\infty t\tau(t)f(t)\exp(-t^2)dt, \tag{26}$$

$$u_r=\xi\frac{4\kappa\Gamma_m}{\sqrt{\pi}}\int_0^\infty\frac{t\psi(t)}{\nu_n(t)}\exp(-t^2)dt, \tag{27}$$

$$u_{lr}=\xi\frac{2\kappa\Gamma_m}{\sqrt{\pi}}\int_0^\infty t\tau(t)[\psi(t)+2J(t)]\exp(-t^2)dt, \tag{28}$$

$$u_a=-\xi\frac{2\kappa\Gamma_m}{\sqrt{\pi}}\int_0^\infty\frac{2\alpha_n(t)+q(t)}{\nu_n(t)}J(t)\exp(-t^2)dt. \tag{29}$$

Here we have introduced the functions

$$\psi(t)=\arctan\frac{t+x}{y}+\arctan\frac{t-x}{y},$$

$$f(t)=x\psi(t)+\frac{y}{2}\ln\frac{y^2+(t-x)^2}{y^2+(t+x)^2},$$

$$J(t)=(x^2-y^2)\psi(t)+2yt+xy\,\ln\frac{y^2+(t-x)^2}{y^2+(t+x)^2} \tag{30}$$

of the dimensionless velocity $t=v/\overline{v}$ and the notation

$$\kappa=\frac{2|G|^2}{\Gamma_mk\overline{v}}, \quad y=\frac{\Gamma(v)}{k\overline{v}}, \quad x=\frac{\Omega(v)}{k\overline{v}}, \tag{31}$$

where κ is the saturation parameter, and x is the dimensionless detuning of the radiation frequency.

According to (26)–(30), u_l is, as expected,[4–11] an odd function of x, and u_r, u_{lr}, and u_a are even functions of x.

5. TRANSPORT COLLISION RATE

To calculate the partial drift velocities (26)–(29), we must know how the transport collision rate depends on the parameter t. For a power potential describing the interaction between the particles,

$$U_i(r)\propto r^{-n}, \tag{32}$$

this dependence can be calculated explicitly:[14]

$$\nu_i(t)=\nu_i(0)\,{}_1F_1(a-1;5/2;-\beta t^2), \quad a\equiv\frac{2}{n}+\frac{1}{2}, \tag{33}$$

where $\beta=M_b/M$ is the mass ratio of the particles of the buffer and absorbing gases, and ${}_1F_1(a;b;x)$ is Kummer's confluent hypergeometric function. It thus follows that the transport collision rate decreases with increasing t for $n<4$ and increases with t for $n>4$. At $n=4$ the collision rate is velocity-independent: $\nu_i(t)=\nu_i(0)=$const. For light buffer particles $(\beta\ll1)$, the $\nu_i(t)$-vs-t dependence in (26)–(29) can

be ignored. As the parameter β grows, the $\nu_t(t)$-vs-t dependence becomes more prominent and reaches a maximum in the case of heavy buffer particles ($\beta \gg 1$).

The partial drift velocity u_a in (29) is caused by the dependence of the transport transition rate on velocity and appears only when $\alpha_i \not\equiv 0$. For the power potential (32),

$$\alpha_i(t) = 2\beta t \frac{n-4}{5n} \frac{{}_1F_1(a;7/2;-\beta t^2)}{{}_1F_1(a-1;5/2;-\beta t^2)}. \tag{34}$$

The function $\alpha_i(t)$ depends most strongly on t when the buffer particles are heavy, just as in the case of $\nu_i(t)$.

6. ANALYSIS OF RESULTS

To specify in Eq. (23) the anomalous dependence of the drift velocity represented by the term u_a, we examine the case where

$$|\nu_m(v) - \nu_n(v)| \ll [\Gamma_m + \nu_m(v)] \frac{\xi}{v}. \tag{35}$$

Here $|u_l|$, $|u_{lr}| \ll u_r$, and the total drift velocity is represented only by the ''ordinary'' (u_r) and anomalous (u_a) reactions of the medium to spontaneous light pressure:

$$u = u_r + u_a. \tag{36}$$

In addition, for simplicity, we assume that the buffer gas pressure is low:

$$\nu_i(v) \ll \Gamma_m. \tag{37}$$

This makes it possible to ignore the dependence of the parameters x and y given by (31) and simplifies Eq. (29) for u_a, since we can now set $q(t) = 0$.

Since the function $\alpha_i(t)$ assumes its maximum value when the buffer gas particles are heavy, the anomalous reaction of the medium, u_a, to light pressure is at its maximum in the $\beta \to \infty$ limit, corresponding to the Lorentz gas model (see, e.g., Ref. 15). For a Lorentz gas in the Doppler broadening limit ($y \to 0$) and with a power potential (32), we can use (27) and (29) to obtain the following expressions for the partial drift velocities if the conditions (35) and (37) are satisfied:

$$\frac{u_r(x)}{u(0)} = 1 - \frac{|x|^{2a}}{\Gamma(a+1)} \exp(-x^2)\, {}_1F_1(1;a+1;x^2),$$

$$\frac{u_a(x)}{u(0)} = 2x^2 - \frac{2|x|^{2a}}{\Gamma(a)} \exp(-x^2)\, {}_1F_1(1;a;x^2), \tag{38}$$

where the parameter a is defined in (33). Here for the arguments of $u_r(x)$ and $u_a(x)$ we take the dimensionless detuning x of the radiation frequency, and the values of $u_r(x)$ and $u_a(x)$ are normalized to the value of the drift velocity $u(x) = u_r(x) + u_a(x)$ at the point $x = 0$.

For small detunings ($x^2 \ll 1$), from (38) we obtain

$$\frac{u_r(x)}{u(0)} = 1 - \frac{|x|^{2a}}{\Gamma(a+1)}, \quad \frac{u_a(x)}{u(0)} = \begin{cases} 2x^2, & n<4, \\ -\dfrac{2|x|^{2a}}{\Gamma(a)}, & n>4. \end{cases} \tag{39}$$

For large detunings ($x^2 \gg 1$), we obtain from (38) the expressions

$$\frac{u_r(x)}{u(0)} = \frac{|x|^{2a-2}}{\Gamma(a)} \exp(-x^2), \quad \frac{u_a(x)}{u_r(x)} = \frac{4-n}{n}. \tag{40}$$

From (38) it follows that the anomalous partial drift velocity $u_a(x)$ is comparable in magnitude to the ''ordinary'' reaction of the medium, $u_r(x)$, to light pressure. The velocity $u_a(x)$ is positive [$\mathbf{u}_a(x) = \mathbf{k} u_a(x)/k$ is directed along the radiation's wave vector $\mathbf{k}$] for power potentials with $n<4$ and negative [$\mathbf{u}_a(x)$ and $\mathbf{k}$ point in the opposite directions] for potentials with $n>4$. From the asymptotic behavior of $u_a(x)$ [the second equation in (40)] we conclude that the positive anomalous addition $u_a(x)$ to the drift velocity is at the maximum for the Coulomb potential ($n=1$), while the negative anomalous addition is at the maximum in the model of hard spheres ($n=\infty$), and one should expect that in magnitude the positive addition is several times larger than the negative.

Figures 1–3 illustrate the results of calculations of the partial ($u_r(x)$ and $u_a(x)$) and total [$u(x) = u_r(x) + u_a(x)$] drift velocities by Eqs. (27) and (29) in conditions (35) and (37) for power potentials. Figure 1 shows the sensitivity of the anomalous addition $u_a(x)$ to the mass ratio β of the buffer and resonant particles. In the case of a heavy buffer gas ($\beta \gg 1$) this addition is at the maximum and strongly influences the spectral dependence of the total drift velocity $u(x)$ (Fig. 1a). Comparison of Figs. 1a and 1b shows that in the current problem the Lorentz gas limit ($\beta \to \infty$) is actually reached at $\beta \approx 3$. In other words, $\beta = 3$ is equivalent to $\beta \gg 1$.

For the Coulomb potential ($n=1$), $u_a(x)$ reaches the maximum value at $x_{\max} = 1.4$ and is greater than the maximum value of $u_r(x)$ reached at $x=0$ by a factor of 1.16 (see Fig. 1a). Calculations show that for power potentials with $n=2$ and $n=3$, the drift velocity $u_a(x)$ reaches the maximum (for $y \ll 1$ and $\beta \gg 1$) at $x_{\max} = 0.87$ and $x_{\max} = 0.77$, respectively, and $u_a(x_{\max})/u_r(0)$ is 0.33 and 0.1, respectively. Thus, $u_a(x_{\max})/u_r(0)$ for potentials with different values of n is approximately equal to $(4-n)/3n$ [cf. the asymptotic behavior of $u_a(x)/u_r(x)$ in (40)].

For $n>4$ the anomalous addition changes sign and its effect on the total drift velocity $u(x)$ results only in a slight narrowing of the spectrum of $u(x)$ (Fig. 2).

In the limit of Doppler absorption-line broadening ($y \to 0$), the anomalous addition vanishes at zero detuning, $u_a(0) = 0$ [see Eq. (39) and Figs. 1 and 2]. But if the condition $y \ll 1$ for Doppler broadening is not met, $u_a(0) \neq 0$ (Fig. 3). Calculations show that for homogeneous broadening ($y \gg 1$) the function $u_a(x)$ is bell-shaped. For example, in the case shown in Fig. 3, when $y=1$ is replaced by $y=5$, curve *3* for all practical purposes becomes curve *2*, i.e., $u_a(x) \approx u_r(x)$. This means that in the case of homogeneous broadening of the absorption line, when light charged particles drift in an atmosphere of heavy charged buffer particles ($n=1$ and $\beta \gg 1$), the anomalous response of the medium to light pressure, $u_a(x)$, increases the total drift velocity $u(x)$ by a factor of exactly 2.

The formulas derived in this paper, which describe the drift of resonant particles in a buffer medium due to the action of a traveling monochromatic light wave, can be gen-

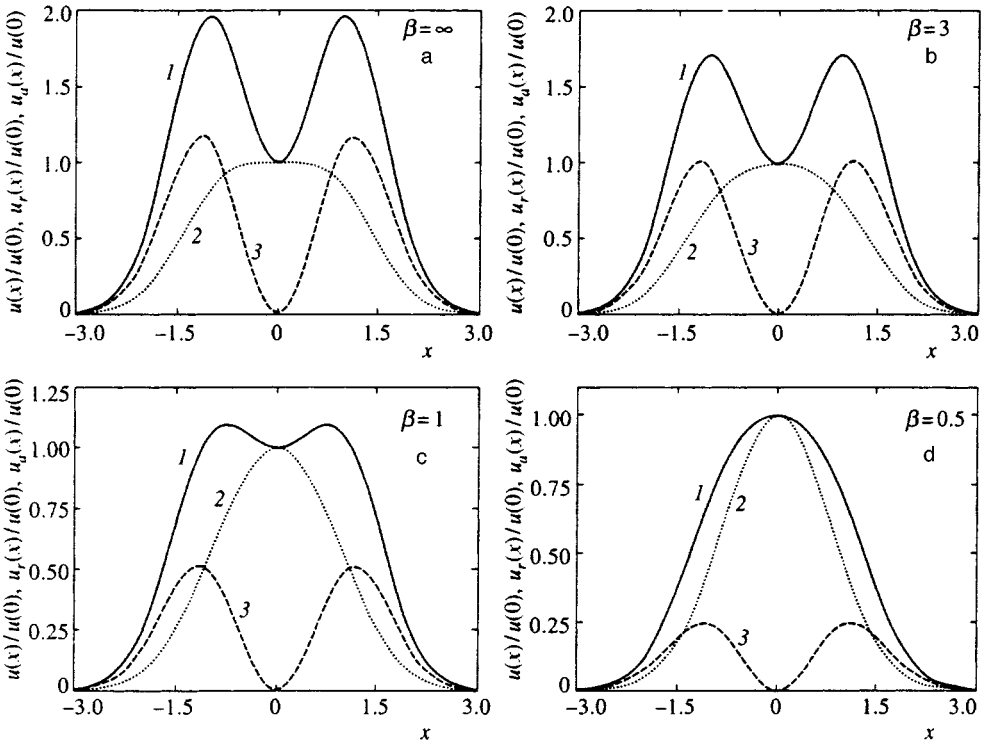

FIG. 1. Dependence of the partial $[u_r(x)$ and $u_a(x)]$ and total $[u(x)]$ drift velocities on the dimensionless detuning x of the radiation frequency for the Coulomb potential $(n=1)$ at different values of the mass ratio $\beta=M_b/M$ of the buffer and resonant particles and $y=10^{-3}$: curves *1* represent $u(x)/u(0)$, curves *2* denote $u_r(x)/u(0)$, and curves *3* represent $u_a(x)/u(0)$.

eralized in a natural way to incorporate the case where broadband radiation of an arbitrary spectral intensity $I(x)$ acts on the medium. Accordingly, we need only replace the saturation parameter κ in Eqs. (26)–(29) by a frequency-dependent parameter $\kappa(x)$,

$$\kappa(x)=\frac{BI(x)}{\pi\Gamma_m k\bar{v}'},\quad B=\frac{\lambda^2\Gamma_m}{4\hbar\omega} \tag{41}$$

(λ is the radiation's wavelength), and then perform additional integration with respect to x over an infinite interval. For instance, when ''white'' light $[I(x)=\text{const}]$ acts on a medium in which the particles interact via the Coulomb potential $(n=1)$, Eqs. (38) imply that $\overline{u_a}/\overline{u_r}=1$ (here the bar on the symbol indicates that the corresponding quantity has been integrated with respect to x). This result suggests, in particular, that allowance for the anomalous response of the medium, $\overline{u_a}$, to light pressure increases the drift velocity $\overline{u}=\overline{u_r}+\overline{u_a}$ by a factor of exactly 2 in the case of light charged particles drifting in an atmosphere of heavy charged buffer particles ($n=1$ and $\beta\gg1$) when the medium is illuminated by ''white'' light with Doppler absorption-line broadening.

In conclusion, we note that our results may be important in examining the role of light pressure in plasmas $(n=1)$ and in studying photocurrent generation in semiconductors caused by photon drag on electrons, especially when electron scattering by charged impurities or defects ($n=1$ and $\beta\gg1$) is predominant, with the effect of transformation of the drift velocity spectrum being at its maximum.

We wish to thank F. Kh. Gel'mukhanov for assistance and interest in this study. We also thank A. M. Shalagin for useful remarks. This work was made possible by financial support from the Russian Fund for Fundamental Research (Grant No. 96-02-19556) and the Netherlands Organization for Scientific Research.

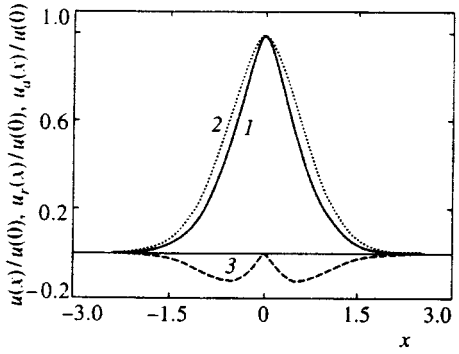

FIG. 2. The same as in Fig. 1 but for $n=8$, $y\to0$, and $\beta\to\infty$.

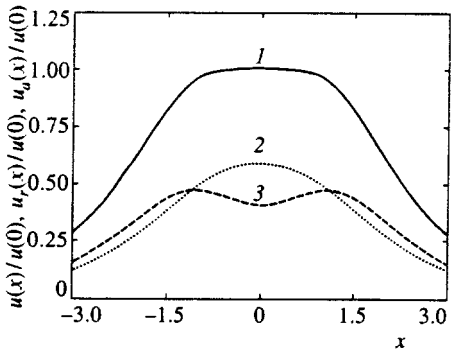

FIG. 3. The same as in Fig. 1 but for $n=1$, $y=1$, and $\beta\to\infty$.

[1)]In the earlier papers,[4,9] the effect was called light-induced diffusion. Dykhne[10] proposed a more appropriate name, light-induced drift, which later became the customary term.

[1] A. P. Kazantsev, Usp. Fiz. Nauk. **124**, 113 (1978) [Sov. Phys. Usp. **21**, 58 (1978).]
[2] V. G. Minogin and V. S. Letokhov, *Laser Light Pressure on Atoms*, Gordon and Breach, New York (1987).
[3] A. P. Kazantsev, G. I. Surdutovich, and V. P. Yakovlev, *Mechanical Effect of Light on Atoms*, World Scientific, Singapore (1990).
[4] F. Kh. Gel'mukhanov and A. M. Shalagin, JETP Lett. **29**, 711 (1979).
[5] G. Nienhuis, Phys. Rep. **138**, 151 (1986).
[6] H. G. C. Werij and J. P. Woerdman, Phys. Rep. **169**, 145 (1988).
[7] S. G. Rautian and A. M. Shalagin, *Kinetic Problems of Non-linear Spectroscopy*, North-Holland, Amsterdam (1991).
[8] E. R. Eliel, Adv. At. Mol. Phys. **30**, 199 (1992).
[9] F. Kh. Gel'mukhanov and A. M. Shalagin, Zh. Éksp. Teor. Fiz. **78**, 1674 (1980) [Sov. Phys. JETP **51**, 839 (1980)].
[10] A. M. Dykhne and A. N. Starostin, Zh. Eksp. Teor. Fiz. **79**, 1210 (1980) [Sov. Phys. JETP **52**, 2380 (1980)].
[11] *The Physics Encyclopedia* [in Russian], A. M. Prokhorov (ed.), Vol. 4, Bol'shaya Rosiĭskaya Éntsiklopediya, Moscow (1994), p. 468.
[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] F. Kh. Gel'mukhanov, Kvant. Élektr. (Moscow) **8**, 1881 (1981) [Sov. J. Quantum Electron. **11**, 1138 (1984)].
[14] F. Kh. Gel'mukhanov, L. V. Il'ichov, and A. M. Shalagin, Physica A **137**, 502 (1986).
[15] E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics*, Pergamon Press, Oxford (1981).

Translated by Eugene Yankovsky
Edited by S. J. Amoretty

Coherent population trapping in an ensemble of three-level atoms in the presence of cooperative relaxation

B. G. Matisov, I. A. Grigorenko

St. Petersburg State Technical University, 195251 St. Petersburg, Russia

I. E. Mazets

A. F. Ioffe Physicotechical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted 11 April 1997)
Zh. Éksp. Teor. Fiz. **112**, 869–876 (September 1997)

We derive equations that describe the dynamics of three-level atoms with a cascade system of levels interacting with two resonant coherent fields in conditions where cooperative relaxation dominates over incoherent spontaneous emission. We calculate the temporal dynamics of the values of the atomic populations. The possibility of coherent population trapping in the presence of the cooperative decay is established. Finally, we calculate the averaged population of the intermediate level as a function of detuning for different values of the coupling constant. [S1063-7761(97)00809-3]

1. INTRODUCTION

Research into the interaction of coherent electromagnetic radiation and multilevel quantum systems is a branch of nonlinear laser spectroscopy and quantum optics that has been rapidly developing. What is important in multilevel systems is that there can be several channels of excitation and induction by laser fields of coherences between long-lived quantum states. This leads to various quantum effects in the inner dynamics of the atoms. The interference of optical channels that emerges because of coherent excitation, which leads in turn to coherent population trapping, forms the basis for the development of many branches of modern physics, such as laser supercooling of atoms and the design of inversion-free lasers (for more details see the reviews in Refs. 1 and 2).

As is known,[1,3] the presence of various relaxation relaxation, say collisional processes, greatly affects the evolution of the populations of atomic systems in coherent population trapping. However, we know of no studies related to coherent population trapping in atomic systems in which $n\lambda^3 \geqslant 1$, where n is the atomic density, and λ is the wavelength of an optical transition. In such conditions the presence of cooperative effects has a marked influence on the evolution of the system.[4,5] What is important in these effects is the self-consistency in the behavior of separate atoms. This leads to a situation in which the dynamics of the quantum system is determined primarily by cooperative relaxation.

In this paper we study the effect of coherent population trapping in an ensemble of three-level atoms with a cascade system of levels interacting with a coherent bichromatic electromagnetic field in the presence of cooperative relaxation.

2. EQUATIONS FOR THE DENSITY MATRIX

Let us state the main approximations used in our theory. We assume that a one-dimensional and homogeneous ensemble of atoms interacts with two resonant coherent electromagnetic fields. We denote the total number of atoms by N, and the atomic states are $|1\rangle_l$, $|2\rangle_l$, and $|3\rangle_l$, where $l=1,\ldots,N$ is the number assigned to an atom.

The orthonormalization condition for the wave function $|i\rangle_l$ $(i=1,2,3)$ has the form

$$ {}_l\langle i|i'\rangle_{l'} = \delta_{ll'}\delta_{ii'}\,. \tag{1}$$

We assume that $|1\rangle_l$ is the lower state, $|3\rangle_l$ is the intermediate state, and $|2\rangle_l$ is the upper state in the cascade system. To describe the behavior of the system we will need a collective atomic operator, which we define as follows:

$$ \hat{R}_{ij} = \sum_{l=1}^{N} |j\rangle_{ll}\langle i|. \tag{2}$$

The commutation rules for $\hat{R}_{ij}$ follow directly from the definition (2):

$$ [\hat{R}_{\mu\nu}, \hat{R}_{\nu'\mu'}] = \delta_{\nu\nu'}\hat{R}_{\mu\mu'} - \delta_{\mu\mu'}\hat{R}_{\nu'\nu}, \tag{3}$$

with $\mu,\ \nu,\ \nu',\mu' = 1, 2, 3$.

The electromagnetic field is described by the Bose operators of photon creation and annihilation, $\hat{a}_j^\dagger$ and $\hat{a}_j$. Then the Hamiltonian describing the interaction of the atoms and field in the rotating wave approximation has the following form:

$$ \hat{H} = \hbar(-\Omega_1\hat{R}_{11} + \Omega_2\hat{R}_{22}) + g_1\hat{a}_1\hat{R}_{31} + g_2^*\hat{a}_2^\dagger\hat{R}_{32} + g_1^*\hat{a}_1^\dagger\hat{R}_{13} + g_2\hat{a}_2\hat{R}_{23}. \tag{4}$$

Here Ω_j is the detuning of the jth field $(j=1,2)$ from the frequency ω_j of the corresponding atomic transition, and

$$ g_j = \sqrt{\frac{2\pi\hbar\omega_j}{V}}\, d_j $$

is the atomic-system-field coupling constant,[4] where ω_1 and d_1 are the frequency and dipole moment of the $|3\rangle$–$|1\rangle$ transition, ω_2 and d_2 are the frequency and dipole moment of the $|2\rangle$–$|3\rangle$ transition, and V is the quantization volume.

To make matters simpler when we describe relaxation processes, we ignore the incoherent decay of the upper lev-

 1063-7761/97/090469-05$10.00

els, which is a consequence of the interaction with the continuum of vacuum (zero-point) modes. This approximation is justified because the relaxation process controlled by the cooperative effect occurs much faster than the incoherent process. Instead of dealing with slow atomic relaxation we introduce fast relaxation of the electromagnetic field to the steady state determined by the laser radiation. Indeed, a pair of laser beams directed at the cell with the gas create a purely coherent state $|\alpha_1\alpha_2\rangle_f$, i.e.,

$$\hat{a}_j|\alpha_1\alpha_2\rangle_f=\alpha_j|\alpha_1\alpha_2\rangle_f, \tag{5}$$

where the α_j are complex numbers. This state persists for the time it takes a photon to pass through the medium with linear dimensions L. The time $\tau=L/c$ can be taken as the shortest time interval for the given system (say, at $L=0.3$ m we have $\tau=1$ ns, which is much shorter than 10^{-7} s, the spontaneous relaxation time). Thus, the field density matrix reaches its steady state $|\alpha_1\alpha_2\rangle_f\,{}_f\langle\alpha_1\alpha_2|$ with a characteristic relaxation time τ.

For the density matrix $\hat{\sigma}(t)$ that fully describes the behavior of the system consisting of the field and the atoms we can write the following expression:

$$\frac{\partial}{\partial t}\,\hat{\sigma}=-\frac{i}{\hbar}[\hat{H},\hat{\sigma}]-\frac{1}{\tau}(\hat{\sigma}-\hat{\sigma}'), \tag{6}$$

where $\hat{\sigma}'$ is the equilibrium density matrix,

$$\hat{\sigma}'=\hat{\rho}\otimes|\alpha_1\alpha_2\rangle_f\,{}_f\langle\alpha_1\alpha_2|, \tag{7}$$

and the reduced density matrix $\hat{\rho}$ describes only the atomic degrees of freedom:

$$\hat{\rho}=\mathrm{Tr}\{\hat{\sigma}\}_f$$

(the trace is taken with respect to the field variables). Equation (6) has a formal solution in the form of a series:[4]

$$\hat{\sigma}=\hat{\sigma}'+\tau\hat{K}_1\hat{\sigma}'+\tau^2\hat{K}_2\hat{\sigma}'+\cdots, \tag{8}$$

in which we leave only the first two terms. Plugging (8) into (6), we obtain the explicit form of the operator $\hat{K}_1$:

$$\hat{K}_1(\hat{X})=-\frac{\partial}{\partial t}\,\hat{X}-\frac{i}{\hbar}[\hat{H},\hat{X}].$$

Thus, to within τ^2 we can write the following expression for $\hat{\sigma}$:

$$\hat{\sigma}=\hat{\sigma}'-\tau\frac{\partial}{\partial t}\,\hat{\sigma}'-\frac{i\tau}{\hbar}[\hat{H},\hat{\sigma}']. \tag{9}$$

Perturbation theory works if $\tau\ll\sqrt{N}g_j/\hbar$ holds, i.e., the photons leave the system faster than the interatomic correlations have time to set in. This means that field variations over time intervals much longer than τ can be excluded adiabatically.

Taking the trace of both sides of Eq. (6) and substituting (9), we arrive at the following equation for the atomic density matrix:

$$\frac{\partial}{\partial t}\,\hat{\rho}-\frac{i\tau}{\hbar}\left[\hat{H}_{\mathrm{at}},\frac{\partial}{\partial t}\,\hat{\rho}\right]$$

$$=-\frac{i}{\hbar}[\hat{H}_{\mathrm{at}},\hat{\rho}]-\frac{\tau}{\hbar^2}\,\mathrm{Tr}\{[\hat{H},[\hat{H},\hat{\rho}]]\}_f, \tag{10}$$

where the Hamiltonian $\hat{H}_{\mathrm{at}}$, which allows only for the atomic variables, has the form

$$\hat{H}_{\mathrm{at}}={}_f\langle\alpha_1\alpha_2|\hat{H}|\alpha_1\alpha_2\rangle_f=\hbar(-\Omega_1\hat{R}_{11}+\Omega_2\hat{R}_{22})$$
$$+g_1\alpha_1\hat{R}_{31}+g_2^*\alpha_2^*\hat{R}_{32}+g_1^*\alpha_1^*\hat{R}_{13}+g_2\alpha_2\hat{R}_{23}. \tag{11}$$

Next we introduce the operator

$$\hat{Q}(\hat{X})=\hat{X}+\frac{i\tau}{\hbar}[\hat{H}_{\mathrm{at}},\hat{X}],$$

and apply it to both sides of Eq. (10). If we ignore terms proportional to τ^2, we get

$$\frac{\partial}{\partial t}\,\hat{\rho}=-\frac{i}{\hbar}\,[\hat{H}_{\mathrm{at}},\hat{\rho}]-\frac{\tau}{\hbar^2}((\langle\hat{H}^2\rangle-\hat{H}_{\mathrm{at}}^2)\hat{\rho}+\hat{\rho}(\langle\hat{H}^2\rangle-\hat{H}_{\mathrm{at}}^2))$$
$$+\frac{\tau}{\hbar^2}(2\hat{H}_{\mathrm{at}}\hat{\rho}\hat{H}_{\mathrm{at}}-2\mathrm{Tr}\{\hat{H}(\hat{\rho}\otimes|\alpha_1\alpha_2\rangle_f\,{}_f\langle\alpha_1\alpha_2|)\hat{H}\}), \tag{12}$$

where $\langle\hat{H}^2\rangle={}_f\langle\alpha_1\alpha_2|\hat{H}^2|\alpha_1\alpha_2\rangle_f$. Then, substituting (4) and (11) in (12), we obtain an equation for the density matrix that describes the ensemble of atoms with a cascade system of levels:

$$\frac{\partial}{\partial t}\,\hat{\rho}=-\frac{i}{\hbar}[\hat{H}_{\mathrm{at}},\hat{\rho}]+\frac{\tau|g_1|^2}{\hbar^2}([\hat{R}_{13},\hat{\rho}\hat{R}_{31}]+[\hat{R}_{13}\hat{\rho},\hat{R}_{31}])$$
$$+\frac{\tau|g_2|^2}{\hbar^2}([\hat{R}_{32},\hat{\rho}\hat{R}_{23}]+[\hat{R}_{32}\hat{\rho},\hat{R}_{23}]). \tag{13}$$

This is a generalization of the well-known Bonifacio model.[6]

3. ATOMIC OPERATORS AND THEIR AVERAGED VALUES

Exhaustive information about the dynamics of an atomic system is contained in the averaged value of the collective atomic operator,[4]

$$r_{ij}\equiv\frac{1}{N}\langle\hat{R}_{ij}\rangle=\frac{1}{N}\,\mathrm{Tr}\{\hat{R}_{ij}\hat{\rho}\},\quad i,\,j=1,\,2,\,3. \tag{14}$$

We use Eq. (13) and write the corresponding equation for r_{ij}:

$$N\,\frac{\partial}{\partial t}r_{ij}=-\frac{i}{\hbar}\,\mathrm{Tr}\{[\hat{H}_{\mathrm{at}},\hat{\rho}]\hat{R}_{ij}\}+\frac{\tau}{\hbar^2}(|g_1|^2$$
$$\times\mathrm{Tr}\{([\hat{R}_{13},\hat{\rho}\hat{R}_{31}]+[\hat{R}_{13}\hat{\rho},\hat{R}_{31}])\hat{R}_{ij}\})+\frac{\tau}{\hbar^2}(|g_2|^2$$
$$\times\mathrm{Tr}\{([\hat{R}_{32},\hat{\rho}\hat{R}_{23}]+[\hat{R}_{32}\hat{\rho},\hat{R}_{23}])\hat{R}_{ij}\}). \tag{15}$$

Employing a cyclic permutation before the trace operation is applied, we get

$$N\,\frac{\partial}{\partial t}r_{ij}=-\frac{i}{\hbar}\,\mathrm{Tr}\{\hat{\rho}[\hat{H}_{\mathrm{at}},\hat{R}_{ij}]\}+\frac{\tau}{\hbar^2}(|g_1|^2$$
$$\times\mathrm{Tr}\{\hat{\rho}\hat{R}_{31}[\hat{R}_{ij'},\hat{R}_{13}]+\hat{R}_{13}[\hat{R}_{31},\hat{R}_{ij}]\hat{\rho}\})$$

$$+\frac{\tau}{\hbar^2}(|g_2|^2\mathrm{Tr}\{\hat{\rho}\hat{R}_{23}[\hat{R}_{ij'},\hat{R}_{32}]+\hat{R}_{32}[\hat{R}_{23},\hat{R}_{ij}]\hat{\rho}\}).$$

Now we can employ the commutation relation (3), which lowers the order by $\hat{R}_{ij}$ on the right-hand side and use the semiclassical approximation[5]

$$\langle\hat{R}_{ij}\hat{R}_{i'j'}\rangle\approx\langle\hat{R}_{ij}\rangle\langle\hat{R}_{i'j'}\rangle, \tag{16}$$

to arrive at the following system of equations:

$$\begin{aligned}
&\dot{r}_{11}=iV_1r_{31}-iV_1^*r_{13}+2\Gamma_1N|r_{31}|^2,\\
&\dot{r}_{22}=iV_2^*r_{32}-iV_2r_{23}-2\Gamma_2N|r_{32}|^2,\\
&r_{11}+r_{22}+r_{33}=1,\\
&\dot{r}_{31}=i\Omega_1r_{31}-iV_1^*(r_{33}-r_{11})+iV_2r_{21}+\Gamma_1Nr_{31}\\
&\qquad\times(r_{33}-r_{11})+\Gamma_2Nr_{32}r_{21},\\
&\dot{r}_{32}=-i\Omega_2r_{32}+iV_1^*r_{12}-iV_2(r_{33}-r_{22})-\Gamma_1Nr_{31}r_{12}\\
&\qquad+\Gamma_2Nr_{32}(r_{22}-r_{33}),\\
&\dot{r}_{21}=i(\Omega_1+\Omega_2)r_{21}+iV_2^*r_{31}-iV_1^*r_{23}\\
&\qquad+(\Gamma_1-\Gamma_2)Nr_{23}r_{31},\\
&r_{ij}=r_{ji}^*,
\end{aligned} \tag{17}$$

where we have introduced the following notation:

$$\Gamma_j=\frac{\tau}{\hbar^2}|g_j|^2,\quad V_j=\frac{g_j\alpha_j}{\hbar}. \tag{18}$$

The system of equations (17) describes the dynamics of the atomic populations r_{ii}. A characteristic feature of this system is the nonlinearity of the terms proportional to Γ_jN, the terms responsible for cooperative relaxation.

For the system of equations (17), for the initial condition we take the equality

$$r_{ij}|_{t=0}=\delta_{i1}\delta_{j1}, \tag{19}$$

which means that all atoms at $t=0$ are in the lowest state $|1\rangle$.

4. RESULTS OF NUMERICAL CALCULATIONS

Below we give the results of numerical integration of the system of equations (17) with the initial conditions (19). We assume for simplicity that $V_1=V_2\equiv V$ and $\Omega_1=-\Omega_2\equiv\Omega$. The calculations were done for various values of Ω, V, and the ratio Γ_2/Γ_1.

Figure 1 depicts the temporal evolution of the atomic populations in an ensemble of atoms with a cascade system of levels. We see that the populations of the lower and upper levels tend to become equal and finite, while the population of the intermediate level tends to zero. A situation like this corresponds to the onset of coherent population trapping in the system. We also note that the steady-state solution $r_{11}=0.5$, $r_{22}=0.5$, and $r_{12}=-0.5$ is an exact solution of the system of equations (17), provided that all the other r_{ij} are zero.

Now look at Fig. 2. For large detunings ($\Omega/\Gamma_1N\gg1$) and coupling constants ($V/\Gamma_1N\gg1$) there can be no coherent population trapping. The steady-state mode is represented by oscillations of the populations of the lower and upper levels, while the population of the intermediate level tends to become finite and constant.

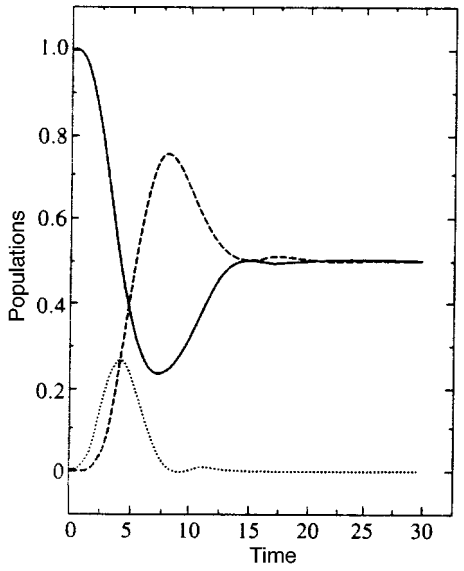

FIG. 1. Temporal evolution of populations in the cascade system at $\Gamma_2/\Gamma_1=10^{-1}$, $\Omega=0$, and $V=0.3\Gamma_1N$. The dimensionless time t/Γ_1N is plotted on the horizontal axis. The solid curve corresponds to the lower level, the dashed curve to the upper level, and the dotted curve to the intermediate level in the cascade system.

Figures 3 and 4 show how the parameter Γ_2/Γ_1 affects the onset of coherent population trapping. For $\Gamma_2/\Gamma_1>1$ (Fig. 4), coherent population trapping becomes impossible, since the state $|2\rangle$ decays at a rate Γ_2N, while the state $|1\rangle$ is populated at a rate Γ_1N, which means that the rate of population of a trapped state is lower than the rate of decay of that state. Figure 5 depicts the suppression of oscillations for a relative large coupling constant ($V>\Gamma_1N$).

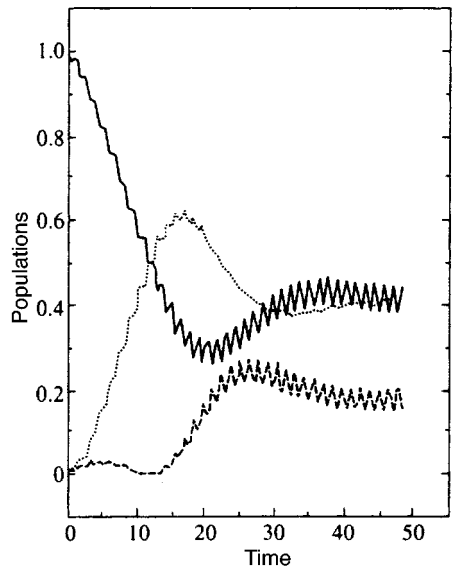

FIG. 2. The same as in Fig. 1 but at $\Gamma_2/\Gamma_1=10^{-1}$, $\Omega=2\Gamma_1N$, and $V=3\Gamma_1N$.

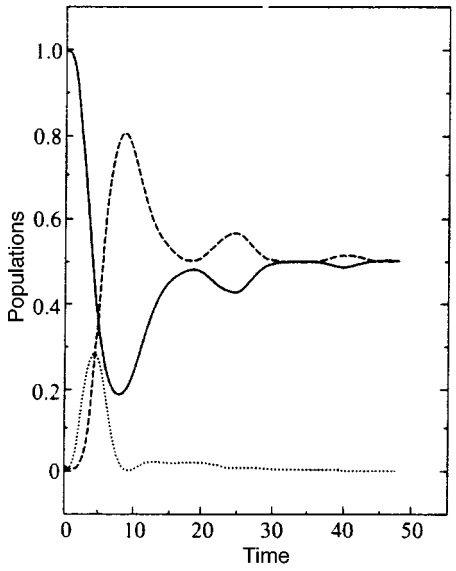

FIG. 3. The same as in Fig. 1 but at $\Gamma_2/\Gamma_1=0.8$, $\Omega=0$, and $V=3\Gamma_1 N$.

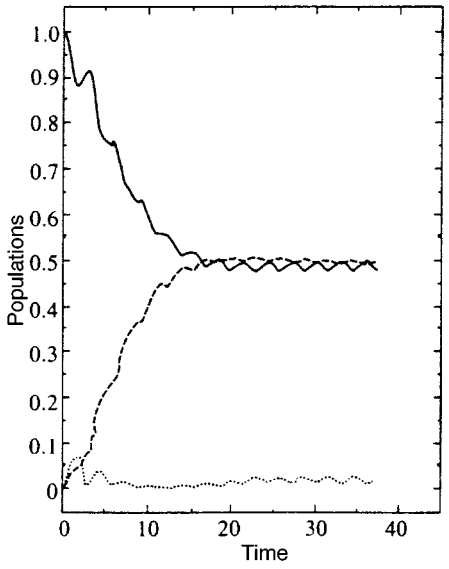

FIG. 5. The same as in Fig. 1 but for $\Gamma_2/\Gamma_1=10^{-1}$, $\Omega=0.3\Gamma_1 N$, and $V=1.5\Gamma_1 N$.

Figure 6 depicts the dependence of the averaged value of the population of the intermediate state $|3\rangle$ on the relative detuning $\Omega/\Gamma_1 N$ for different values of the parameter V characterizing the intensity of the laser field. We see that the width of the black line (the gap in the fluorescence spectrum; see Ref. 1) increases with the coupling constant.

Analysis of the numerical solutions of the system of equations (17) suggests that in the case of a cascade system of levels, coherent population trapping is established in the presence of cooperative relaxation only if

$$\frac{\Gamma_2}{\Gamma_1}\ll 1, \quad V\gg\Gamma_2 N.$$

The first condition coincides with, while second differs considerably from, the condition for the onset of coherent population trapping for a similar level diagram in the presence of coherent relaxation.[1] Moreover, when cooperative dynamics is present, population oscillations begin to play an important role in the initial stages of evolution, provided that $V\ll\Gamma_2 N$. But if $V\gg\Gamma_2 N$, the oscillations are suppressed. On the other hand, if incoherent relaxation is predominant, the oscillations are suppressed at Rabi frequencies higher than the rate of relaxation from the upper level, and are well-developed in the opposite case.[7]

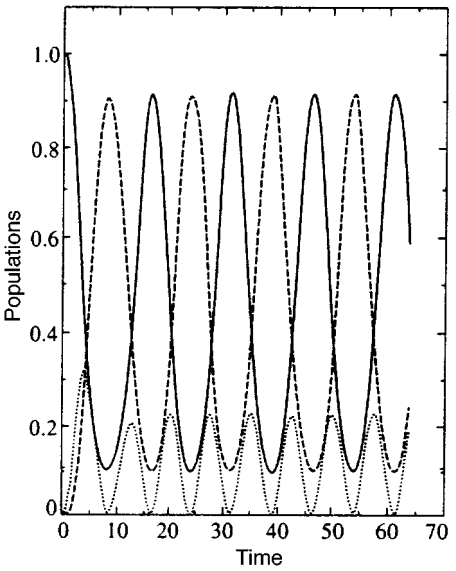

FIG. 4. The same as in Fig. 1 but at $\Gamma_2/\Gamma_1=2$, $\Omega=0$, and $V=3\Gamma_1 N$.

5. CONCLUSION

We have done a theoretical study of the establishment of coherent population trapping in an atomic ensemble interact-

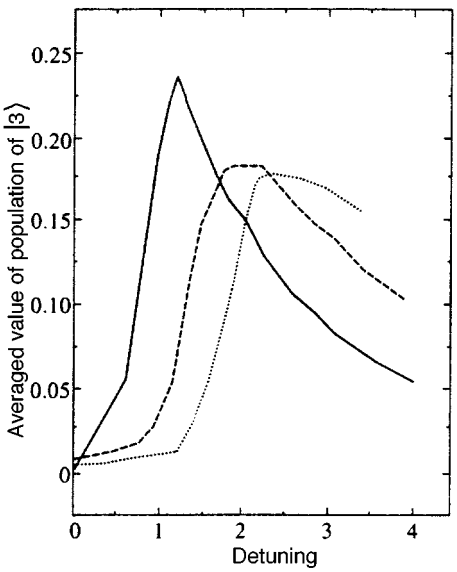

FIG. 6. Dependence of the averaged value of the population of the intermediate state $|3\rangle$ on the dimensionless detuning $\Omega/\Gamma_1 N$ for different values of the parameter V at $\Gamma_2/\Gamma_1=10^{-1}$. The solid curve corresponds to $V=0.5\Gamma_1 N$, the dashed curve to $V=1\Gamma_1 N$, and the dotted curve to $V=2\Gamma_1 N$.

ing with two resonant coherent electromagnetic fields in the presence of cooperative relaxation. The behavior of this quantum system is nonlinear because there is a nonlinear relaxation process. The results make it possible to use coherent population trapping for, say, coherent bleaching,[8] in optical media with an average distance between the atoms of order of the wavelength of an atomic transition, $n\lambda^3 \sim 1$. We have found that cooperative relaxation has a strong effect on the dynamics of the atomic ensemble. We have also established the properties of this dynamics, as compared to the case where incoherent spontaneous relaxation dominates, and the conditions under which coherent population trapping is established in the system examined.

This work was sponsored by the State Committee of the Russian Federation for Higher Learning (Grant No. 5-5.5-139).

[1] B. D. Agap'ev, M. B. Gornyĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Usp. Fiz. Nauk **163**, No. 9, 1 (1993) [Phys. Usp. **36**, 763 (1993)].
[2] E. Arimondo, in *Progress in Optics*, Vol. 35, E. Wolf (ed.), North-Holland, Amsterdam (1995), p. 256.
[3] E. Arimondo, Phys. Rev. A **54**, 2216 (1996).
[4] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics: Superradiation, Bistability, and Phase Transitions* [in Russian], Nauka, Moscow (1988).
[5] M. M. Al'perin, Ya. D. Kubis, and A. N. Khizhnyak, *Introduction to the Physics of Two-Level Systems* [in Russian], Naukova Dumka, Kiev (1987).
[6] R. Bonifacio, P. Schwendimann, and F. Haake, Phys. Rev. A **4**, 302 (1971).
[7] E. A. Korsunskiĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Zh. Éksp. Teor. Fiz. **102**, 1096 (1992) [Sov. Phys. JETP **75**, 595 (1992)].
[8] M. B. Gornyĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Zh. Éksp. Teor. Fiz. **95**, 1263 (1989) [Sov. Phys. JETP **68**, 728 (1989)]; K. J. Boller, A. Imamoglu, and S. E. Harris, Phys. Rev. Lett. **66**, 2593 (1991).

Translated by Eugene Yankovsky

Convective model of a microwave discharge in a gas at atmospheric pressure in the form of a spatially localized plasma

A. A. Skovoroda

Russian Scientific Center ''Kurchatov Institute,'' 123182 Moscow, Russia
(Submitted 21 January 1997)
Zh. Éksp. Teor. Fiz. **112**, 877–893 (September 1997)

Experiments and a theoretical model consistent with them are presented which show that a stationary microwave discharge in a gas at atmospheric pressure under the action of free convection due to the action of the buoyant force on the heated air can be spatially localized, taking a spheroidal shape. Vortex motion inside the spheroid gives this localized plasma formation some of the properties of a material body which are manifested in a distinct material isolation from the surrounding space, in the formation of a narrow thermal boundary layer and flow separation, and in the formation of secondary vortices in the wake region. The characteristic radius of the stationary localized plasma is governed mainly by the wavelength of the microwave radiation $a \sim 0.137\lambda$. Energy balance is established to a significant degree by convective cooling of the microwave-heated structure. © *1997 American Institute of Physics.* [S1063-7761(97)00909-8]

1. INTRODUCTION

Microwave discharges in a gas at atmospheric pressure are well known.[1,2] They have been studied and widely used for a long time, for example, in the construction of plasmotrons for various technological needs.[3]

Plasma in microwave discharges is practically in equilibrium, and the roughly identical temperature for all charged and neutral components does not exceed $T=1$ eV. A large number of models of equilibrium microwave discharges are known which quite accurately predict the temperature.[2–5] Raĭzer[2] notes that the exponential dependence of the conductivity σ (the density n_p) of the plasma on T makes the temperature insensitive to important details of the microwave discharge model. All models of an equilibrium microwave discharge explain the presence of a threshold power, below which its stationary existence is impossible. The magnitude of the threshold power can depend substantially on the model.

Any model of an equilibrium microwave discharge contains the solutions of two interrelated problems, electrodynamic and thermal. The solution of the electrodynamic problem determines the heating power of the electromagnetic waves; the solution of the thermal power determines their cooling power (heat conduction and convection); the power balance determines the equilibrium temperature of the plasma (gas) and the threshold microwave power. Microwave discharge models differ as to their geometry (planar, cylindrical, spherical), scheme for microwave power feed (waveguide, resonator, free space), how they take account of motion of the gas (forced or free convection), and the degree of detail in the solution of the electrodynamic and thermal problems. A distinguishing feature of all models is the necessity of calculating the coefficient of reflection (absorption) of the electromagnetic wave by the plasma. An increase in the reflection (decrease in absorption) of the microwave power with growth of the temperature explains the experimental fact of the relative smallness of T in microwave discharges in comparison with other types of equilibrium discharges (arc, inductive HF, optical).[2]

Let us briefly characterize the known models. The overwhelming majority of the models are constructed in cylindrical (planar) geometry[1)] for dimensions of the plasma comparable with the dimensions of the discharge chamber. In such models Maxwell's equations are solved together with the equations of heat conduction with boundary conditions at the walls bounding the plasma. As a rule, the wall temperature and incident microwave power flux (the wall is assumed to be transparent to microwaves) are fixed. To obtain analytical solutions, the approximation of sharp plasma boundary (the analog of the channel model for an arc) is frequently used.[2] Exact solutions of Maxwell's equations and the equation of heat conduction with possible account of forced convection of the gas are obtained numerically.[3–5] These models adequately describe actual experimental conditions in commercially available plasmotrons.

It is known from experiment that microwave discharge plasmas have spatial structure.[1,3,6–9] The traditional models described above have been successfully applied to explain the characteristic features of those structures that are essentially related, first, to the dimensions of the discharge chamber, second, to the distribution of electromagnetic waves in the chamber, and third, to the nature of induced convection of the gas in the chamber.[3,10] However, experiments indicate that a microwave discharge in free space (in particular, in large discharge chambers) also have spatial structure.[6–9,11,12] The characteristic features of the structures arising in them are not associated with the position of any walls. Attempts have been made to use this property of electrode-free microwave discharges in air to explain the phenomenon of ball lightning.[13–15]

A spatially localized plasma of spheroidal shape in free space is observed in experiments with microwave beam focusing.[6–9] In such experiments the location of the plasma is determined by the position of the focus, and its characteristic dimensions are determined by the diffraction-

limited dimensions of the neck of the microwave beam at the focus. The well known model of such a microwave discharge is similar in its main features to the model of an optical discharge at the focus of a laser beam.[2] If the beam is pointed upward the speed of propagation of the ionization front V_n (Ref. 16) (analog of the well-known normal velocity of slow combustion) downward from the focus (breakdown region) is compensated by the rise of the hot gas under the action of the repulsive buoyancy force. The attainment of stationary equilibrium is facilitated by the decrease of V_n with distance from the focus.[8]

Experiments show that a localized plasma of spheroidal shape in a microwave discharge in free space is formed and exists as a stationary object without focusing of electromagnetic waves.[11,12,17] The question arises, what determines the size of the localized plasma in this case? Timofeev[5] has attempted to answer this question within the framework of the traditional model (see above) without account of the motion of the gas. For spherically symmetric microwave heating, the wall of the discharge chamber may be placed at a distance substantially exceeding the characteristic dimensions of the localized plasma. Such a situation can arise in experiments with microwave discharges in low-density gases in metal discharge chambers whose dimensions significantly exceed one wavelength.

Experimentally, however, a localized plasma of spheroidal shape may arise arise and stably exist in air at atmospheric pressure for asymmetric (even from one side) power feed (see below). The present paper is dedicated to a study of this aspect of microwave discharges, based on the inclusion of natural air convection.

The paper is organized as follows. Section 2 describes the experimental basis of the model. An analysis of the experimental data allows us to conclude that the localized plasma formation is materially isolated from the surrounding air. This is due to the vortical nature of the motion of the hot air inside the formation. Subsection 2 A contains a brief description of two experiments carried out by the author together with V. A. Zhil'tsov especially to provide an experimental basis for the described convective model. In what follows all quantitative estimates relate to these two experiments. Section 3 formulates the main points of the convective model. The model is reduced to its simplest form to make it possible to obtain analytical dependences. Section 4 discusses the model of a microwave discharge of spheroidal shape on the basis of the relations obtained in Section 3 and calculates the threshold power flux and the size and temperature of the localized plasma. These quantities are compared with the experimental data. This section analyzes the need for further improvements and numerical experiments. The Conclusion summarizes the main conclusions of this work.

2. EXPERIMENTAL BASIS OF THE CONVECTIVE MODEL

We begin this section with a description of experiments which were performed especially to provide an experimental basis against which to check the convective model. They are characterized by asymmetric unfocused feed of the microwave power to the localized plasmas.

Figure 1 shows two improvements of the setup described

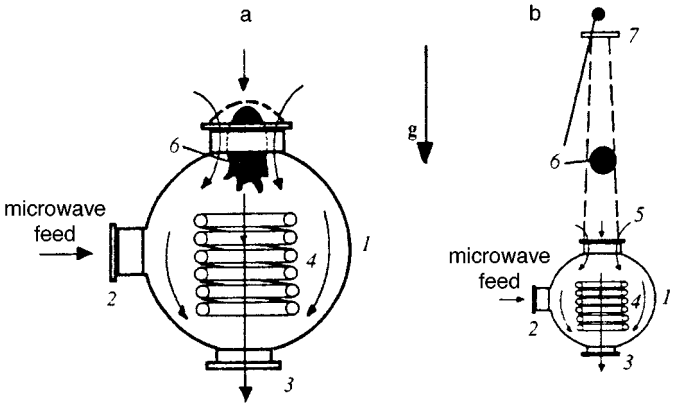

FIG. 1. Experimental setups for demonstrating an immobile (a) and mobile (b) localized plasma formation: *1*—discharge chamber, *2*—port with SHF power feed, *3*—evacuation port, *4*—discharge stabilizing inductor, *5*—upper port with brass gauze shaped into a hemisphere (a) or conical waveguide (b), *6*—localized plasma, *7*—upper port of the gauze waveguide. The arrows indicate the motion of air, the direction of the force of gravity and power feed.

in detail in Refs. 12 and 17 (frequency of the stationary magnetron 2.45 GHz, maximum power 5 kW, diameter of the cylindrical copper discharge chamber 16 cm, length 38 cm). The improvements pertain only to the instrumentation located on the upper port of the discharge chamber. In the case corresponding to Fig. 1a, the opening of the port is covered by braided brass gauze (netting), pushed upward into a hemispherical shape (cell dimensions 3×3 mm^2, wire diameter 0.5 mm). In the case corresponding to Fig. 1b, a conical waveguide 60 cm in length and made from the same gauze, is fastened to the upper port. The entrance diameter of the waveguide is 10 cm and its (open) exit diameter is 7 cm.

Figures 2 and 3 show photographs of a stationary discharge created in the setup depicted in Fig. 1a. By regulating the air flow rate V_∞ through the lower port (in the region of the upper flange $V_\infty \sim 0.5-2$ m/s) and the microwave power flux S through the side flange ($S\sim 10-30$ W/cm^2 in the region of the localized plasma formation) it is possible to achieve stable heating of a spheroidal discharge lasting many hours (characteristic radius of exposure $a\sim 2-4$ cm) separated from the walls. In Fig. 1a the arrows pointing downward show the trajectories of the smoke jets which visualize the flows of cold air around the plasma. Evaporation of ceramic probes inside the localized plasma formation visualized the flows of hot air upward inside the discharge $V_b\sim 1-3$ m/s. Rapid burn-up of the particles and strong distortion of the character of the motion inside the plasma resulting from the introduction of the ceramic probes hindered visualization. However, the vortical nature of this motion was clearly revealed (see also Ref. 17).

In Figs. 2 and 3 it can be seen that the upper, frontal part of the localized plasma is always stable and is distinctly isolated from the cold air. Infrared photographs confirm the presence of a sharp boundary between the hot plasma and the cold air (the characteristic dimension of the transition layer is $\delta\sim 0.5$ cm). The lower, aft part of the formation is always unstable (see, e.g., the view from above in Fig. 3b), and the region of hot air forms a protracted tail, trailing off into the

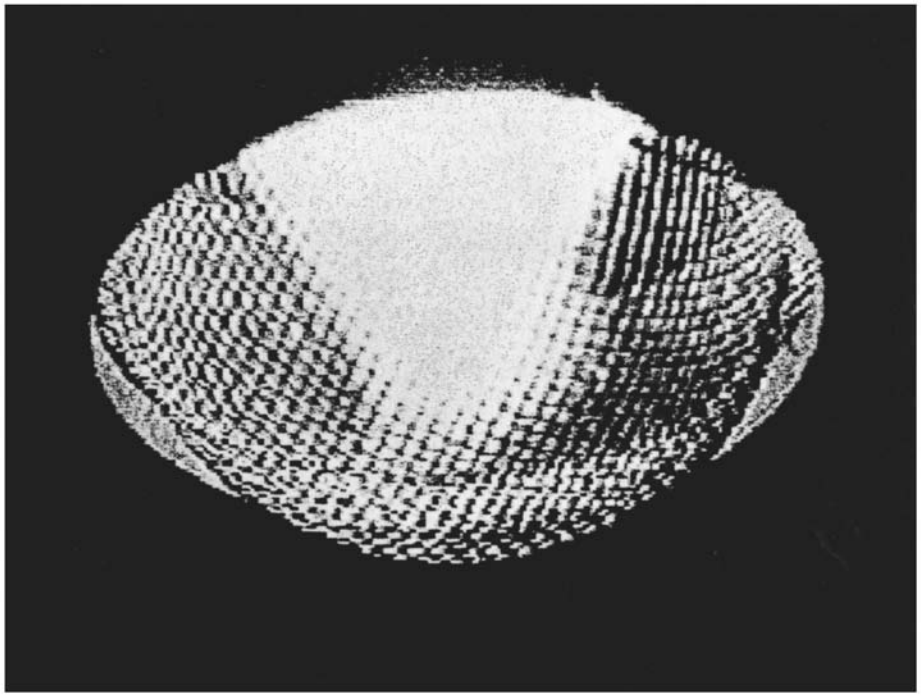

FIG. 2. Photograph of a stationary discharge in the setup depicted in Fig. 1a. The upper port, overhead shot.

evacuation system. The transition from stable to unstable flow is observed in the region of the equator of the localized plasma.

We performed a few simple experiments which indicated a certain degree of material isolation of the localized plasma formation from the surrounding medium. When argon was let into the chamber outside the plasma, Ar lines were not observed in the optical spectrum, whereas after brief evaporation of NaCl inside the localized plasma the characteristic sodium line was observed for an extended time. When the microwave discharge was blasted with the flame from a gas torch, the flame enveloped the plasma, leaving it intact.

Figure 4 presents photographs of a moving localized plasma taken at two times in the setup shown in Fig. 1b. The appearance of the localized plasma was initiated by a sharp metal wire below the gauze waveguide, and the spheroidal object moved upward along the gauze waveguide, all the while keeping its shape, and finally leaving it at the other end. These two times are also recorded in Fig. 4. The upward motion took place with variable velocity in the range $V_\infty \sim 0.5-2$ m/s (it varied along the waveguide; this was apparently due to variation of the microwave matching of the waveguide with the generator during the movement of the plasma). To demonstrate the degree of isolation of the localized plasma, we performed the following simple experiment. From above the waveguide we installed a glass plate in its path, cooled to liquid-nitrogen temperature. On it we detected a number of little metal spheres (of the material of the wire initiating the discharge) 0.03–0.07 mm in diameter. This indicates that the metal evaporated from the wire during initiation of the discharge was transported a substantial distance by the localized plasma.[2)]

The facts indicating the material isolation of localized hot plasmas have been known for a long time. In fact it was just these facts that initiated the discussion: does the gas flow through or around the localized plasma?[2] Gus'kov *et al.*[18] showed that the observed speed of a slowly moving optical

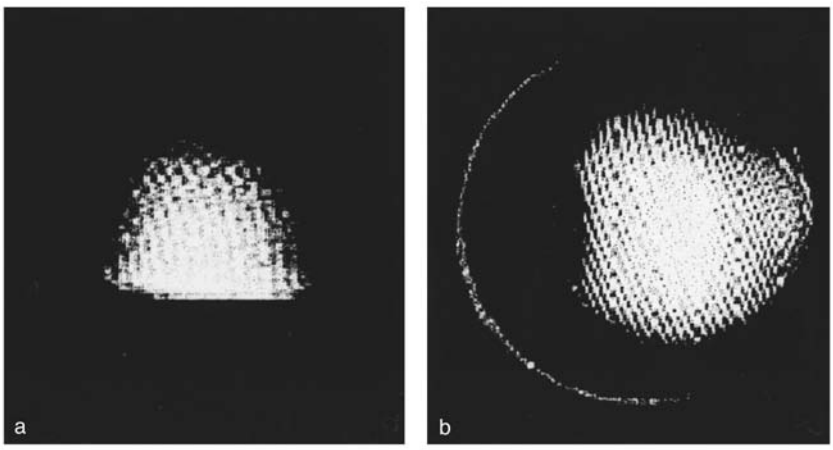

FIG. 3. Photographs of a stationary discharge in the setup depicted in Fig. 1a: a—side shot at the level of the upper port, b—overhead shot.

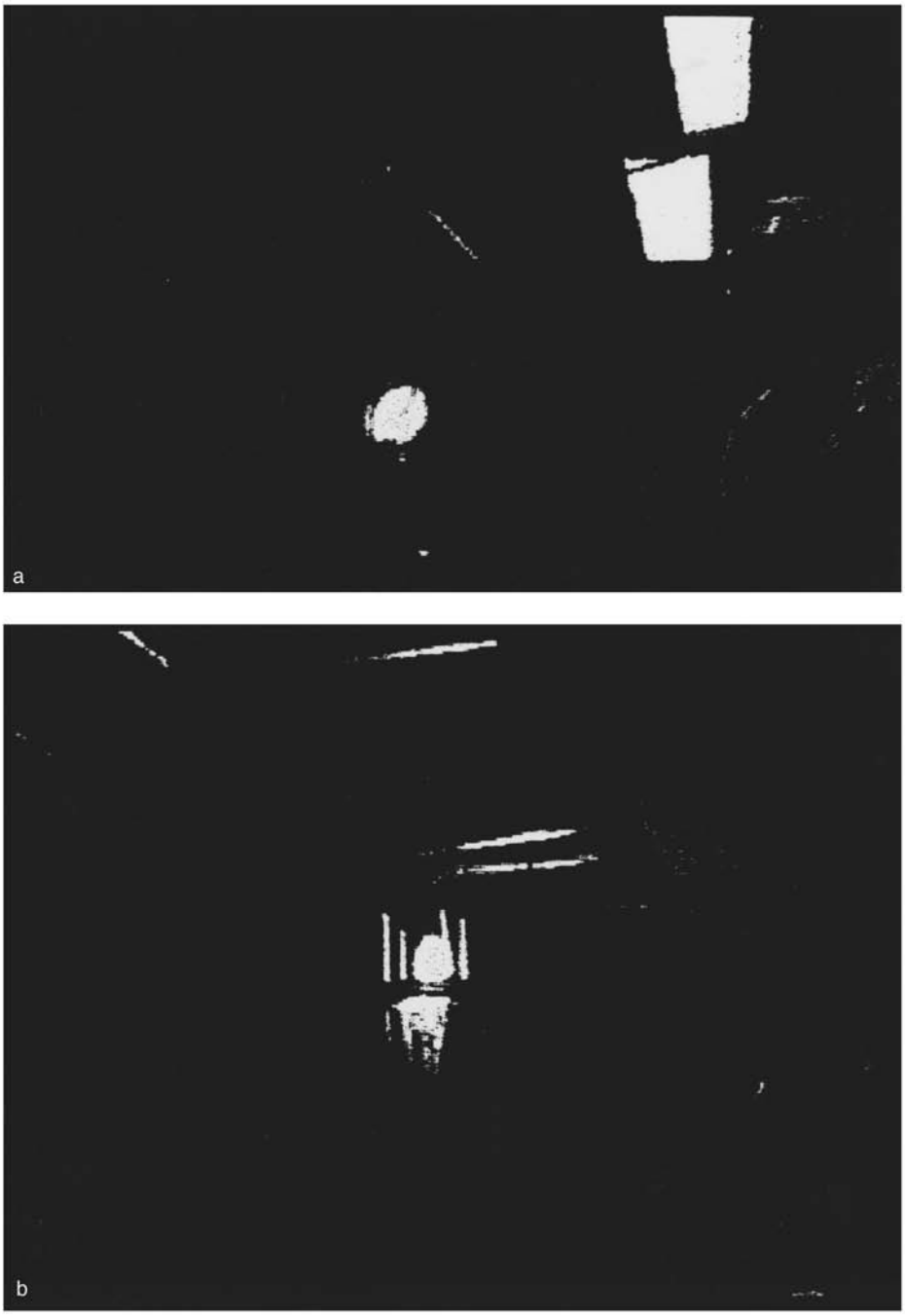

FIG. 4. Photographs of a mobile discharge in the setup depicted in Fig. 1b: a—at the time when the localized plasma formation is located in the middle of the waveguide (on the right features of the experimenter and lamp are visible); b—at the time when the formation is passing through the upper port of the gauze waveguide (the ends of the four mounting rods of the waveguide are visible).

discharge can be approximated by the formula

$$V_{\propto} \approx V_n \sqrt{\frac{T_0}{T_e}}, \tag{1}$$

where T_e is the temperature of the surrounding medium and T_0 is the temperature at the center of the discharge, proceeding from the model of flow of a dense cold gas around a sphere of hot, rarefied gas. Raĭzer[2] notes (see p. 411) that ''...about 10% of the advancing gas flows through the discharge, the remaining 90% flows around the heated region as though around a solid body.'' The low chemical efficiency of microwave plasmotrons noted in Refs. 3 and 4 is due to this effect.

The physical reason for the material isolation and shape stability of the localized plasma is the vortical motion inside the hot region. The vortical nature of the motion has been revealed in many calculations of plasmotrons.[3,4,18] The natural reason for circulation of the hot air is free convection. Gorbunov *et al.*[19] calculated the evolution of a sphere of hot air (radius 13 m, initial temperature 2500 K, position above sea level 1800 m) in the gravitational field. Free convection leads to the formation of the observed upward rising toroidal vortex. The natural atmospheric phenomenon of ''thermals'' is explained in a similar manner—a long-lived spheroidal region of low-density hot air.[20] Zhil'tsov *et al.*[17] pointed out the analogy of a microwave discharge in the form of a local-

ized plasma formation with thermals. Dem'yanov and Inogamov carried out calculations analogous to those of Gorbunov *et al.*[19] for the experimental conditions of Zhil'tsov *et al.*[17] (initial radius $a\sim 1.5$ cm, temperature $T_0=4000$ K, at sea level) and demonstrated good agreement with the observed rate of rise of the localized plasma inside a discharge chamber, $V_\infty\sim 0.5$ m/s, and the rate of spreading of the spheroid, 0.1 m/s. In their calculations air flows around the plasma like around a solid body and air turbulence is manifested in the wake region.

3. MAIN POINTS OF THE CONVECTIVE MODEL

The main point of the convective model is the introduction of a spherical surface with radius a, where the normal component of the velocity of the air particles vanishes. This reflects the experimental fact of the material isolation of the localized plasma from the surrounding air. This is a simplification, since it is well known that a small fraction of the air nevertheless passes through the plasma. Concerning how to take this fact into account in the model, see below.

For the most part, we are interested in the outer region. For the outer region, which can have its own complicated ''fine'' structure, we assume a constant temperature $T=T_0$ up to the bounding surface. That the temperature is essentially uniform inside localized plasmas is indicated by infrared photographs. For stationary subsonic motion of a gas inside a region with uniform temperature the continuity condition $\operatorname{div}\mathbf{V}=0$ should be fulfilled. The azimuthally symmetric, axially bounded solution of this equation with zero velocity normal to the spherical boundary surface has the form

$$V_r=C\left(1-\frac{r^2}{a^2}\right)\cos\theta,\quad V_\theta=-C\left(1-\frac{2r^2}{a^2}\right)\sin\theta. \tag{2}$$

Here we have introduced the spherical coordinate system with origin at the center of the spherical surface and the angle θ is measured from the vertical. The constant C will be determined below. Solution (2) describes a Hill vortex and shows that convective heat transfer inside the localized plasma formation can in principle quite effectively equalize the temperature. This may serve as an additional justification of the assumption of uniform temperature inside the region. Plasma is found only inside the sphere. The plasma density is $n_p=\text{const}$ inside the spherical boundary surface and $n_p=0$ outside it. This is the channel model approximation, which is usually invoked (see, e.g., Ref. 2).

A uniform flow of cold air with velocity V_∞ flows around the immobile spherical boundary surface. This flow balances the buoyancy force in the analysis of the experiment with the stationary localized plasma (see Fig. 1a) or appears when one uses a moving coordinate system bound to the moving plasma (see Fig. 1b). It is possible either to specify this velocity from the results of measurements or estimate it from the following arguments. We write out the balance between the buoyancy force and the drag force of a material sphere:

$$\frac{1}{2}\,\zeta\rho_e V_\infty^2 S_p=g(\rho_e-\rho)V_p. \tag{3}$$

Here ρ is the air density in the localized plasma formation, ρ_e is the density of the surrounding air, g is the acceleration due to gravity, V_p is the volume of the formation, S_p is its cross-sectional area, and ζ is the drag coefficient. Expressing the ratio of air densities in terms of the corresponding ratio of temperatures for constant pressure $\rho/\rho_e=T_e/T_0$, we obtain the following expression for the velocity of a spherical localized plasma:

$$V_\infty=\sqrt{\frac{8}{3\zeta}\left(1-\frac{T_e}{T_0}\right)ga}. \tag{4}$$

Treating the localized plasma formation as a material sphere for the Reynolds number corresponding to our experimental conditions $\mathrm{Re}=V_\infty a/\nu_e\approx 1000$, where ν_e is the kinematic viscosity of the surrounding air, we estimate the drag coefficient ζ from the known experimental dependence $\zeta=\zeta(\mathrm{Re})$ (see, e.g., Ref. 21) as $\zeta=0.5$. If we substitute characteristic measured values ($T_0=4000$ K, $T_e=300$ K, $a=3$ cm) into formula (4), we obtain $V_\infty\approx 1.2$ m/s (Ref. 17). This value agrees very closely with the experimentally observed velocity. The above estimates may serve as another clear indication of the material isolation of the localized plasma.

3.1. Heating of a localized plasma by the field of a electromagnetic plane wave

In the model it is assumed that the plasma sphere is heated by a linearly polarized electromagnetic plane wave from one side, namely from below. The classical problem of scattering of an electromagnetic plane wave by a sphere of arbitrary radius with arbitrary electrical properties was solved by Mie in 1908.[22] We will make use of his results represented in convenient form in Ref. 23. If the power flux of the electromagnetic wave S is given, then the rate of heat flow into the sphere W is given by

$$W=(Q_t-Q_s)S, \tag{5}$$

where the cross sections of transmission and scattering $Q_{t,s}$ have the form

$$Q_t=-\frac{\lambda^2}{2\pi}\,\mathrm{Re}\sum_{j=1}^{\infty}(2j+1)(a_j+b_j),$$

$$Q_s=\frac{\lambda^2}{2\pi}\,\mathrm{Re}\sum_{j=1}^{\infty}(2j+1)(|a_j|^2+|b_j|^2). \tag{6}$$

Here λ is the wavelength of the electromagnetic wave in vacuum, and the complex coefficients a_j, b_j are given in the Appendix.

3.2. Convective heat exchange between a localized plasma and the surrounding space

It is well known that at large Reynolds numbers heat exchange between a body and an air flow takes place in the thermal boundary layer, the characteristic magnitude of which $\delta\sim a/\sqrt{\mathrm{Re}}\ll a$ (Ref. 21). In our experiment we also observed such a narrow layer in the front part of the localized plasma. Therefore, it is completely logical to apply the well-developed theory of boundary layers from aerodynam-

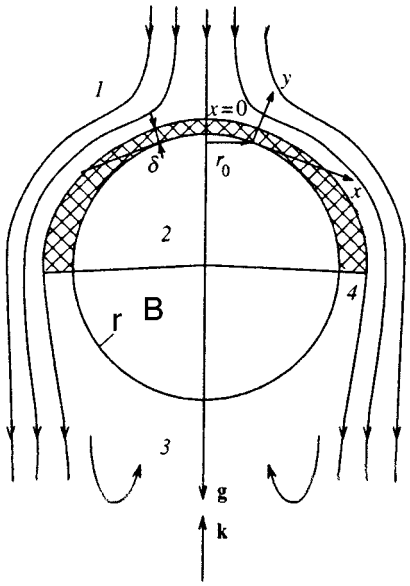

FIG. 5. Basic diagram of streamline air flow around the localized plasma: B—boundary surface, *1*—air flowlines, *2*—localized plasma formation, *3*—swirled wake region, *4*—flow separation region. The thermal boundary layer is hatched. Also shown are the coordinate system employed and directions of the force of gravity and propagation of the electromagnetic plane wave.

ics (see, e.g., Refs. 21 and 25) to our problem. In what follows we will make wide use of the results and notation of Ref. 25.

Introducing the coordinate system shown in Fig. 5, we may write the boundary layer equations in the form (see Appendix A and Ref. 25)

$$\frac{\partial r_0\rho u}{\partial x}+\frac{\partial r_0\rho w}{\partial y}=0,$$

$$u\,\frac{\partial u}{\partial x}+w\,\frac{\partial u}{\partial y}=\frac{\rho_e}{\rho}\,u_e\,\frac{\partial u_e}{\partial x}+\frac{1}{\rho}\,\frac{\partial}{\partial y}\,\mu\,\frac{\partial u}{\partial y}+g_x\left(1-\frac{\rho_e}{\rho}\right),$$

$$c_p\left(u\,\frac{\partial T}{\partial x}+w\,\frac{\partial T}{\partial y}\right)=\frac{1}{\rho}\,\frac{\partial}{\partial y}\,k\,\frac{\partial T}{\partial y}. \tag{7}$$

Here u and w are the velocity components along the x and y axes, g_x is the z component of the force of gravity, k is the thermal conductivity, $\mu=\nu\rho$ is the coefficient of dynamic viscosity, c_p is the specific heat, u_e is the longitudinal velocity on the outer boundary layer, $r_0=a\,\sin(x/a)$, and account is taken of the fact that the thickness of the layer δ is much smaller than the radius of the sphere a.

If we multiply the first of Eqs. (7) by c_pT, the third by r_0, and add them, integrate over y from 0 to $h\gg\delta$ and again make use of the equation of continuity, we obtain the well-known formula for the density P of the heat flux from the surface bounding the localized plasma:[25]

$$P=c_p\rho_e(T_0-T_e)\,\frac{d(r_0u_e\theta_T)}{r_0dx},$$

$$\theta_T(x)=\int_0^\infty\frac{\rho u}{\rho_e u_e}\,\frac{T-T_e}{T_0-T_e}\,dy. \tag{8}$$

The required total power of convective cooling of the localized plasma W_T is obtained by integrating the heat flux density (8) over x with the weighing factor $2\pi r_0$ from 0 to x_a:

$$W_T=2\pi c_p\rho_e(T_0-T_e)(r_0u_e\theta_T)_{x=x_a}. \tag{9}$$

We have limited the integration to the flow separation point $x=x_a$.[21,25] It is well known that in the numerical solution of the boundary layer equations this point corresponds to the time at which the solution fails and the boundary layer equations become unsuitable. The accumulated experience of aerodynamics shows that the flow separation point of a body calculated in this way accurately corresponds to reality. The model assumes that the main heat loss from the localized plasma takes place in the frontal laminar boundary layer (see below).

For simplicity we neglect the effect of the force of gravity on the motion of the gas in the layer, i.e., we neglect the last term on the right-hand side of the second of Eqs. (7). Let us see why such a simplification is possible under our experimental conditions. We first determine for which air flow velocity V_∞^* the first term on the right-hand side of the second of Eqs. (7) is comparable with the last term in the frontal part of the localized plasma. Toward this end, we make use of the well-known result for potential flow around a sphere[21,25]

$$u_e=\frac{3}{2}\,V_\infty\,\sin\left(\frac{x}{a}\right). \tag{10}$$

We obtain the following estimate for V_∞^*:

$$V_\infty^*\approx\sqrt{\frac{4}{9}\left(1-\frac{T_e}{T_0}\right)ga}.$$

This value is substantially smaller than the value obtained using formula (4), which agrees with experiment.

To obtain an analytical expression for θ_T we employ the generalized Dorodnitsyn–Faulkner–Scan–Mangler transformation[25]

$$d\eta=\left(\frac{u_e}{\nu_e\bar{x}}\right)^{1/2}\frac{\rho}{\rho_e}\,\frac{r_0}{a}\,dy,\quad d\bar{x}=\left(\frac{r_0}{a}\right)^2dx,$$

$$\rho r_0u=\frac{\partial\psi}{\partial y},\quad \rho r_0w=-\frac{\partial\psi}{\partial x},\quad \psi=a\rho_e\sqrt{u_e\nu_e\bar{x}}f(\bar{x},\eta).$$

If we introduce the function f in this way, the continuity equation is automatically satisfied and the remaining two equations of Eqs. (7) now read[25]

$$\left(\frac{\mu\rho}{\mu_e\rho_e}\,f''\right)'+m_1ff''+m\left(\frac{\rho_e}{\rho}-f'^2\right)$$
$$=\bar{x}\left(f'\,\frac{\partial f'}{\partial\bar{x}}-f''\,\frac{\partial f}{\partial\bar{x}}\right),$$

$$\left(\frac{\mu\rho}{\mu_e\rho_e\mathrm{Pr}}\,T'\right)'+m_1fT'=\bar{x}\left(f'\,\frac{\partial T'}{\partial\bar{x}}-T'\,\frac{\partial f}{\partial\bar{x}}\right). \tag{11}$$

Here the temperature has been normalized to T_0, the Prandtl number is $\mathrm{Pr}=\mu c_p/k$, $u=u_ef'$, $m_1=(m+1)/2$, $m=(\bar{x}/u_e)(\partial u_e/\partial\bar{x})$, and the primes denote differentiation with respect to η. For air the Prandtl number depends very

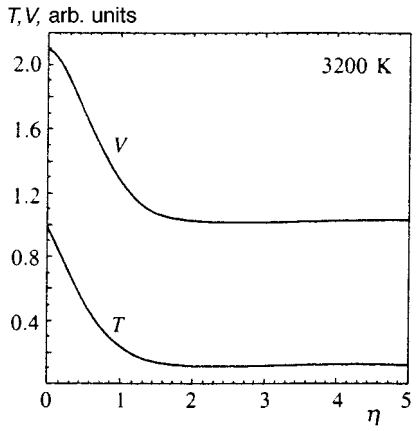

FIG. 6. Result of calculation of the dependence of the dimensionless temperature $T=T/T_0$ and velocity $V=f'/u_e$ on the parameter η (transverse to the layer) for the temperature $T_0=3200$ K of the localized plasma and the temperature $T_e=300$ K of the surrounding air.

weakly on temperature, and we take it to be constant: Pr=0.7. The product of the dynamic viscosity and the air density depends weakly on the temperature, and for the ratio $b=\mu\rho/\mu_e\rho_e$ we use the constant value $b=0.5$, obtained by taking the ratio of the tabulated values at 4000 and 300 K.

At the origin of the boundary layer for $x/a<1$ a good approximation is the solution when f and T are only functions of η (Ref. 25). Expanding in the small parameter x/a, we find that $m=1/3$ and $m_1=2/3$. In this case the solution is found by solving the following system of ordinary differential equations:[25]

$$bf'''+\frac{2}{3}ff''+\frac{1}{3}\left(\frac{T_0}{T_e}T-f'^2\right)=0.$$

$$\frac{b}{\mathrm{Pr}}T''+\frac{2}{3}fT'=0. \tag{12}$$

These equations should be augmented by the boundary conditions. So far, without special discussion, we have used three obvious conditions: $T=1$ for $\eta=0$, $T=T_e/T_0$ and $f'=1$ for $\eta=h\gg 1$. The vanishing of the normal component of the velocity ($w=0$) on the boundary surface ($\eta=0$) leads to another boundary condition $f=0$ for $\eta=0$. As the missing boundary condition we can use the continuity of the longitudinal velocity u and its normal derivative on the boundary surface ($\eta=0$). However, we will use the simpler and physically clearer, but approximate[3)] boundary condition $f''=0$ on the boundary of the localized plasma. Indeed, for motion, for example, in the equatorial plane from inside the plasma toward the boundary the velocity u increases [see formula (2)]; for motion toward the boundary from outside the velocity also increases and not only outside the boundary layer [see formula (10)], but also inside it due to heating of the gas. It is natural to assume that the maximum velocity is reached at the boundary.

Equations (12) can only be solved numerically. Figure 6 provides an example of such a solution which shows that the entire fall of the temperature and velocity is concentrated in the region $\eta<1$, i.e., a boundary layer is actually formed. Table I lists the calculated velocity on the boundary u_0 in units of the velocity outside the layer u_e for different ratios of the temperature outside and inside the spheroid, T_0/T_e. For the conditions characteristic of our experiments, the velocity is accurately approximated on the boundary by the formula $V_i\approx u_e\sqrt[3]{T_0/T_e}V_\infty$ in Eq. (2). For the temperature we may adopt the accurate approximate formula

$$\frac{T}{T_0}\approx 1+\left(\frac{T_e}{T_0}-1\right)\mathrm{erf}\left(\eta\sqrt{\frac{\mathrm{Pr}}{3b}\left(\frac{T_0}{T_e}\right)^{1/3}}\right), \tag{13}$$

where erf is the error function.[24]

For θ_T we obtain the approximate expression

$$\theta_T\approx\frac{1}{r_0}\left(\frac{T_0}{T_e}\right)^{1/6}\sqrt{\frac{b\nu_e x^3}{\pi u_e\mathrm{Pr}}}. \tag{14}$$

TABLE I.

T_0/T_e	u_0/u_e
2	1.178
4	1.457
6	1.683
8	1.877
10.67	2.106
12	2.205
16	2.487
20	2.738
40	3.734

4. CONVECTIVE MODEL OF A MICROWAVE DISCHARGE IN THE FORM OF A LOCALIZED PLASMA FORMATION

Let us analyze the relations obtained in the previous section. To start with, let us consider heating. Figure 7a shows how the microwave power absorption cross section (5) normalized to the cross-sectional area of the sphere, $a=(Q_t-Q_s)/\pi a^2$, varies as a function of the radius of the sphere under our experimental conditions at fixed temperature. A distinct maximum around $a\sim 1.7$ cm can clearly be seen. For this radius the ratio of the perimeter of the equatorial cross section of the sphere to the wavelength is equal to 0.86:

$$\frac{2\pi a}{\lambda}\sim 0.86. \tag{15}$$

This is the well-known condition for the fundamental electrical resonance of a sphere.[23,26]

Let us now consider aspects of convective cooling. A number of observations and calculations of laminary boundary layers on material spheres show that the flow separation point is located near the equator. This does not contradict our experimental observations. Taking $x_a=\pi a/2$ in our estimates of the convective cooling power in expression (9), we obtain

$$W_T\approx\pi a^2 c_p\rho_e V_\infty(T_0-T_e)\xi,$$

$$\xi=\frac{\sqrt{3}\pi}{2}\left(\frac{T_0}{T_e}\right)^{1/6}\sqrt{\frac{b}{\mathrm{Pr\ Re}}}. \tag{16}$$

The balance of heating and cooling power is expressed as

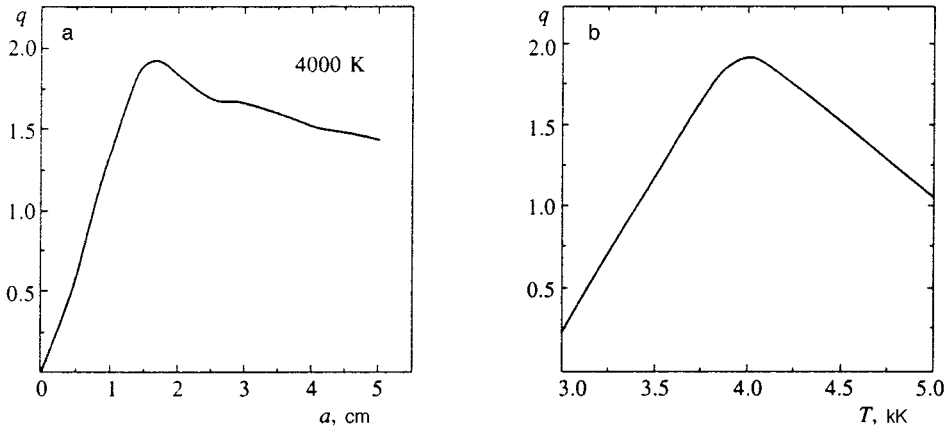

FIG. 7. Dependence of the microwave absorption cross section q normalized to πa^2: a—on the radius a of the localized plasma when the plasma temperature is 4000 K; b—on the temperature of the localized plasma (maximum value). Wavelength 12 cm.

$$q(a,T_0)S=\xi c_p\rho_e V_\infty(T_0-T_e). \tag{17}$$

With decrease of the microwave power flux S Eq. (17) can hold up to the point at which the absolute maximum of the surface $q(a,T_0)$ touches the aT_0 plane.[4)] Thus, near its disappearance threshold the localized plasma formation should have the diameter determined by the radiation wavelength as given by relation (15).

For our experimental conditions we obtain $a\sim 1.7$ cm, $T_0\sim 4000$ K, and threshold power flux $S\sim 20$ w/cm^2. For the characteristic thickness of the transition layer we obtain the estimate

$$\frac{\delta}{a}\sim\frac{T_0}{T_e}\frac{1}{\sqrt{\mathrm{Re}}}\approx 0.2-0.3.$$

These values are found to be in good agreement with the experimental values described above (see Sec. 2).[5)]

4.1. Discussion

The above estimates show that the convective model describes quite accurately the main features of a microwave discharge in the form of a localized plasma. The characteristic dimension of such plasmas is governed by the radiation wavelength through relation (15). This state of affairs was indeed predicted by P. L. Kapitsa in 1955.[13] The larger the microwave frequency, the smaller the dimensions and higher the temperature of the localized plasma. Its stable shape is tied up with the formation of the primary vortex of the plasma under the action of free convection, the gas in which is subject to the greatest heating by the microwaves, with the result that the plasma density in it is maximum. Vortex motion inside the localized plasma gives rise to a boundary layer which is opaque for the surrounding gas. In the hydrodynamic sense the localized plasma acquires the properties of a material body which are manifested in the formation of a boundary layer, in flow separation from it, and in the presence of a swirled (secondary vortices) wake region (see Fig. 5).

The formation of secondary vortices explains the observed instability of the wake region of the localized plasma and the magnitude of the drag coefficient ζ in Eq. (3). In the model described here this region was in fact excluded from consideration. However, the aft region of the plasma turns out to be of fundamental importance, in particular, for understanding how some quantity of the surrounding gas penetrates into its center. Indeed, while the upper, frontal critical point of the localized plasma (at which the velocity vanishes) is stable (the rising flow of hot air encounters the opposing flow of cold air caused by the external pump or by the motion of the plasma itself), this cannot be said of the lower, aft critical point (the flow in the secondary vortices which is due to the rise of air heated in the wake region can be in the same direction as inside the plasma).

What process stabilizes the position of the plasma boundary around the aft critical point? In this region an ionization front propagates downward with normal velocity V_n counter to the flow.[16] This situation was in fact considered in Ref. 18 Let us make some rough estimates, following the logic of Refs. 2 and 18. From the condition for continuity of the flow entering the localized plasma from its aft region in the region of the axis, $\rho_e V_n=\rho V$, and relations (2) we obtain the estimate

$$V_\infty\sim V_n\left(\frac{T_0}{T_e}\right)^{2/3},$$

which differs only slightly from Eq. (1). The fraction of gas passing through the localized plasma is roughly estimated by the ratio V_n/V_∞ and is quite small.[2] If we express the normal velocity of propagation of the front in terms of the energy flux density S_T (Ref. 2) escaping through the front, $V_n\sim S_T/c_p\rho_e T_0$, and assume that the wake region is in contact with the localized plasma over an area on the order of πa^2, then the conductive cooling power in the wake region $W_k=\pi a^2 S_T$ can be of the same order of magnitude as the convective cooling in the layer W_T (16) (the parameter ξ under our conditions is of the same order of magnitude as $(T_e/T_0)^{2/3}$). These estimates do not take account of the fact that the gas in the wake region has been heated up by the convective heat flux emanating from the boundary layer. Including this would lower the conductive energy losses from the primary vortex, which in this case are substantial, in the overall energy balance.

Necessary improvements in the model can go in different interrelated directions. On the one hand, remaining within the framework of the principles expounded above, it

is necessary to calculate the position of the flow separation point with allowance for the effect of the force of gravity in the layer and the actual temperature dependence of the various parameters involved (viscosity, specific heat, etc.) under the conditions of the localized plasma. Other boundary conditions can also be considered. On the other hand, it is most likely that in its internal structure the localized plasma is closer to a toroidal vortex that entrains part of the air from the aft region in the region of the axis. The entrained air is heated inside the plasma and leaves it by way of the upper critical point. Under these conditions a ''sublayer problem'' can be formulated, where the properties of this sublayer are assigned by the axial flow of entrained air.[25] The model should also include a description of the wake region.

5. CONCLUSION

A stationary microwave discharge in a gas at atmospheric pressure under the action of free convection due to the action of the buoyancy force on the heated air can be spatially localized, taking a spheroidal shape. Vortical motion inside the spheroid gives this localized plasma some of the properties of a material body, which are manifested in a certain degree of material isolation that it has from the surrounding space, in the formation of a narrow thermal boundary layer and flow separation, and in the formation of secondary vortices in the wake region.

The characteristic radius of the localized plasma is determined by the wavelength of the microwave radiation, $a \sim 0.137\lambda$. Convective heat exchange plays an important role in the energy balance of localized plasmas.

ACKNOWLEDGMENTS

I thank V. A. Zhil'tsov and P. M. Kosarev, without whose help the the experiments could not have been carried out, and A. Yu. Dem'yanov and N. A. Inogamov for calculating the temporal dynamics of a spherical volume of hot air. This work was greatly influenced by discussions with V. A. Zhil'tsov, Zh. Lyaĭtner, É. A. Manykin, A. V. Timofeev, and P. H. Handel.

APPENDIX A

1. The coefficients in the Mie formulas have the following form:

$$a_j = -\frac{j_j(N\xi)[\xi j_j(\xi)]' - j_j(\xi)[N\xi j_j(N\xi)]'}{j_j(N\xi)[\xi h_j^{(1)}(\xi)]' - h_j^{(1)}(\xi)[N\xi j_j(N\xi)]'},$$

$$b_j = -\frac{j_j(\xi)[N\xi j_j(N\xi)]' - N^2 j_j(N\xi)[\xi j_j(\xi)]'}{h_j^{(1)}(\xi)[N\xi j_j(N\xi)]' - N^2 j_j(N\xi)[\xi h_j^{(1)}(\xi)]'}.$$

Here j_j and $h_j^{(1)}$ are the spherical Bessel functions of the first and third kinds,[24] the primes denote differentiation with respect to their argument $\xi = 2\pi a/\lambda$, and $N = n + i\chi$ is the complex refractive index of the plasma. For the refractive index we use the usual expressions (see, e.g., Ref. 2)

$$n = \sqrt{\frac{\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2}}, \quad \chi = \sqrt{\frac{-\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2}},$$

$$\varepsilon_1 = 1 - \frac{\omega_p^2}{\omega^2 + \nu_{en}^2}, \quad \varepsilon_2 = \frac{\omega_p^2 \nu_{en}}{\omega(\omega^2 + \nu_{en}^2)},$$

where ω_p is the electron plasma frequency, ν_{en} is the electron collision frequency, and ω is the microwave frequency. In our estimates, we used the following approximations taken from Ref. 5 for the relations linking the plasma density and the collision frequency in air with the temperature of the localized plasma:

$$n_p(T_0) = 5.91 \cdot 10^{15} \exp\left(-\frac{14.42}{T_0 - 1.74}\right),$$

$$\nu_{en}(T_0) = 5.74 \cdot 10^{11} \frac{T_e}{T_0} \sqrt{T_0}.$$

Here the plasma density is in units of cm^{-3}, the collision frequency, s^{-1}, and the temperature, kK.

2. The surface layer approximation consists in making use of the fact that the velocity, temperature, and thickness of the layer vary only slightly in one direction (along the x axis) and strongly in the normal direction (along the y axis). This leads to the natural condition that the pressure p transverse to the layer is constant (we exclude from the pressure its hydrostatic part associated with the potential force due to gravity).[25] The pressure on the outer boundary of the layer (sufficiently removed from the spherical boundary surface introduced in the formulation of the problem) is determined by the Bernoulli equation $u_e^2 + p/\rho_e = \mathrm{const}$, where u_e is the longitudinal velocity on the outer boundary of the layer.[6)] In the derivation of the surface layer equations we made use of the constancy of the specific heat c_p, the absence of viscous heating, and smallness of the Mach number $M_e = u_e/\sqrt{(\gamma-1)c_p T_e} \approx 0.003$ (Ref. 25).

[1)] Spherical geometry was used in Refs. 3 and 5.

[2)] When the moving localized plasma did not form, the small metal spheres were not detected. The material of the gauze waveguide did not evaporate as the plasma formation moved through it.

[3)] In the approximation $\delta \ll a$ the two conditions coincide.

[4)] ξV_∞ depends weakly on temperature and radius.

[5)] Note that in the comparison it is necessary to take two circumstances into account: as a rule, the diameter of the localized plasma on the photographs in visible light is somewhat increased due to the exposure effect, and its measured temperature is somewhat decreased due to averaging over its cross section.

[6)] At sufficient distances from the localized plasma, potential subsonic incompressible inviscid motion of the gas is assumed.

[1] P. L. Kapitsa, Zh. Éksp. Teor. Fiz. **57**, 1801 (1969) [Sov. Phys. JETP **30**, 973 (1970)].

[2] Yu. P. Raĭzer, *Gas Discharge Physics* (Springer, New York, 1991).

[3] *Low-Temperature Plasma: HF and UHF Plasmotrons*, No. 6 [in Russian] (Nauka, Novosibirsk, 1992).

[4] V. M. Lelevkin and D. K. Otorbaev, *Experimental Methods and Theoretical Models in Non-Equilibrium Plasma Physics* (Ilim, Frunze, 1988).

[5] A. V. Timofeev, Fiz. Plazmy **23**, 1 (1997) [Plasma Phys. Rep. **23**, (1997)].

[6] A. L. Vikharev, P. A. Ivanov, and A. G. Litvak, in *Proc. Int. Workshop ''Microwave Plasma and Its Applications''* (Zvenigorod, Russia, 1994), p. 391.

[7] V. G. Brovkin, Yu. F. Kolesnichenko, and D. V. Khmara, in *Ball Lightning in the Laboratory* [in Russian] (Khimiya, Moscow, 1994), p. 119.

[8] A. S. Zarin, A. A. Kuzovnikov, and V. M. Shibkov, *Freely Localized Microwave Discharge in Air* [in Russian] (Neft' i Gaz, Moscow, 1996).

[9] E. T. Protasevich, in *Ball Lightning in the Laboratory* [in Russian] (Khimiya, Moscow, 1994), p. 165.

[10] B. É. Meĭerovich, Zh. Éksp. Teor. Fiz. **61**, 1891 (1971) [Sov. Phys. JETP **34**, 1006 (1972)].
[11] Y. H. Ohtsuki and H. Ofuruton, Nature (London) **350**, 139 (1991).
[12] V. A. Zhil'tsov, A. A. Skovoroda, and A. V. Timofeev, Zh. Éksp. Teor. Fiz. **106**, 1687 (1994) [JETP **79**, 912 (1994)].
[13] P. L. Kapitsa, Dokl. Akad. Nauk SSSR **101**, 245 (1955).
[14] P. H. Handel, in *Science of Ball Lightning (Fire Ball)*, edited by Y. H. Ohtsuki (World Scientific, Singapore, 1989), p. 254.
[15] A. B. Mitsuk and V. E. Mitsuk, in *Ball Lightning in the Laboratory* [in Russian] (Khimiya, Moscow, 1994), p. 151.
[16] Yu. P. Raĭzer, *Laser-Induced Discharge Phenomena* (Consultants Bureau, New York, 1977).
[17] V. A. Zhil'tsov, Zh. F. Lyaĭtner, É. A. Manykin *et al.*, Zh. Éksp. Teor. Fiz. **108**, 1966 (1995) [JETP **81**, 1072 (1995)].
[18] K. G. Gus'kov, Yu. P. Raĭzer, and S. T. Surzhikov, Kvantovaya Elektron. **17**, 937 (1990) [Sov. J. Quantum Electron. **20**, (1990)].
[19] S. Yu. Gorbunov, B. N. Gordeĭchik, A. P. Darintsev *et al.*, Zh. Prikl. Mekh. Tekh. Fiz. **5(195)**, 70 (1992).
[20] N. K. Vinnichenko and N. Z. Pinus, *Turbulence in the Free Atmosphere*, Consultants Bureau, New York (1980).
[21] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon Press, Oxford, 1987).
[22] G. Mie, Ann. Phys. (Leipzig) **25**, 377 (1908).
[23] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).
[24] M. Abramowicz and I. A. Stegun (editors), *Handbook of Mathematical Functions* (Dover, New York, 1965).
[25] T. Cebeci and P. Bradshaw, *Physical and Computational Aspects of Convective Heat Transfer* (Springer-Verlag, 1984).
[26] H. Bateman, *Mathematical Analysis of Electrical and Optical Wave Motion on the Basis of Maxwell's Equations* (Dover, New York, 1955).

Translated by Paul F. Schippnick

Effect of an electron beam generated in an *X*-pinch plasma on the structure of the *K* spectra of multiply charged ions

S. A. Pikuz, T. A. Shelkovenko, and V. M. Ramanova

P. N. Lebedev Physical Institute, Russian Academy of Sciences, 117927 Moscow, Russia

J. Abdallah, Jr., G. Csanak, and R. E. H. Clark

Los Alamos National Laboratory, Los Alamos, New Mexico 87545 USA

A. Ya. Faenov and I. Yu. Skobelev

State Scientific Center ''All-Russian Scientific-Research Institute of Physicotechnical and Radio Engineering Measurements,'' 141570 Mendeleevo, Moscow Province, Russia

D. A. Hammer

Cornell University, Ithaca, New York, 14854 USA
(Submitted 3 February 1997)
Zh. Éksp. Teor. Fiz. **112**, 894–909 (September 1997)

The first experimental studies of an electron beam generated in an *X* pinch on the XP machine (Cornell University, USA) and the BIN machine (P. N. Lebedev Physical Institute, Russian Academy of Sciences) are reported. It is shown that it is possible in an *X* pinch to isolate the effect of a plasma-generated electron beam on the multiply charged ion radiation. The intensities of the satellite lines corresponding to Li-, Be-, B-, and C-like ions are calculated for the Al spectrum on the basis of a collisional–radiative model with a non-Maxwellian electron distribution in the plasma. The effect of an electron beam on the multiply charged light ion radiation in an *X*-pinch plasma is demonstrated. Comparing our calculations with the experimental spectra, we conclude that the present model can be used to estimate the electron beam intensity. © *1997 American Institute of Physics.* [S1063-7761(97)01009-3]

1. INTRODUCTION

The spectra of multiply charged ions, corresponding to electron transitions to the *K* shell (*K* spectra) are as a whole the least complicated and most susceptible to analysis and numerical calculations. The most interesting feature of these spectra is the presence of satellite structures near the lines of the resonance series of H- and He-like ions. The relative intensities of the satellite lines carry information about the plasma parameters and are widely used in plasma studies as a diagnostic.[1] However, this is valid for a stationary equilibrium plasma when along with the resonance lines of H- and He-like ions, only satellites corresponding to transitions of ions with ionization number less by unity are observed. The spectra of He-like ions of elements with $Z \approx 10-20$ observed in a laser plasma heated by neodymium laser radiation with pulse duration $\tau = 1-5$ ns may serve as an example here. In this case, the satellite lines corresponding to transitions of Li-like ions have noticeable intensity (Fig. 1a). Despite the presence of almost ten lines in the satellite structure and two main channels for populating the upper levels (dielectronic recombination and direct excitation by electron impact), such spectra are now uniquely interpreted and serve as a good instrument for measuring the parameters of a plasma.

However, in a number of experiments on the *K* spectra of elements with $Z = 12-15$ much more complicated satellite structures have been observed containing simultaneously the lines of several multiply charged ions. Such spectra were first recorded in a plasma heated by CO_2-laser radiation.[2,3] Groups of lines in the long-wavelength segment of the spectrum have been identified with transitions of Be-...F-like ions, whose presence together with He- and Li-like ions has been explained by the rapid change of the plasma. The presence of many hundreds of near-lying lines, the absence of reliable calculations of the atomic constants for the observed transitions prevents a detailed analysis of the spectra; however, even in the well-studied segment of the spectrum containing satellite lines corresponding to transitions of Li-like ions it has not been possible to describe the intensities of these lines in terms of the time-dependent model. Such spectra have been observed in plasma heated by femtosecond laser radiation[4] (Figs. 1b and 1c). Here also it has not been possible to uniquely describe the line intensities in the satellite structures of the resonance lines of He-like ions without invoking additional ion excitation processes in the plasma.

At the same time, there are experimental data on the presence in plasma heated by CO_2-laser radiation of nanosecond duration as well as by shorter-wavelength laser radiation of femtosecond duration, of high-power electron beams capable of having a substantial effect on excitation processes in the plasma. However, the problem of allowing for the effect of an electron beam on a laser plasma is extraordinarily complicated since it is necessary to simultaneously treat the time ans space varition of the plasma for a spectrum containing several hundred lines. Besides, reliable methods

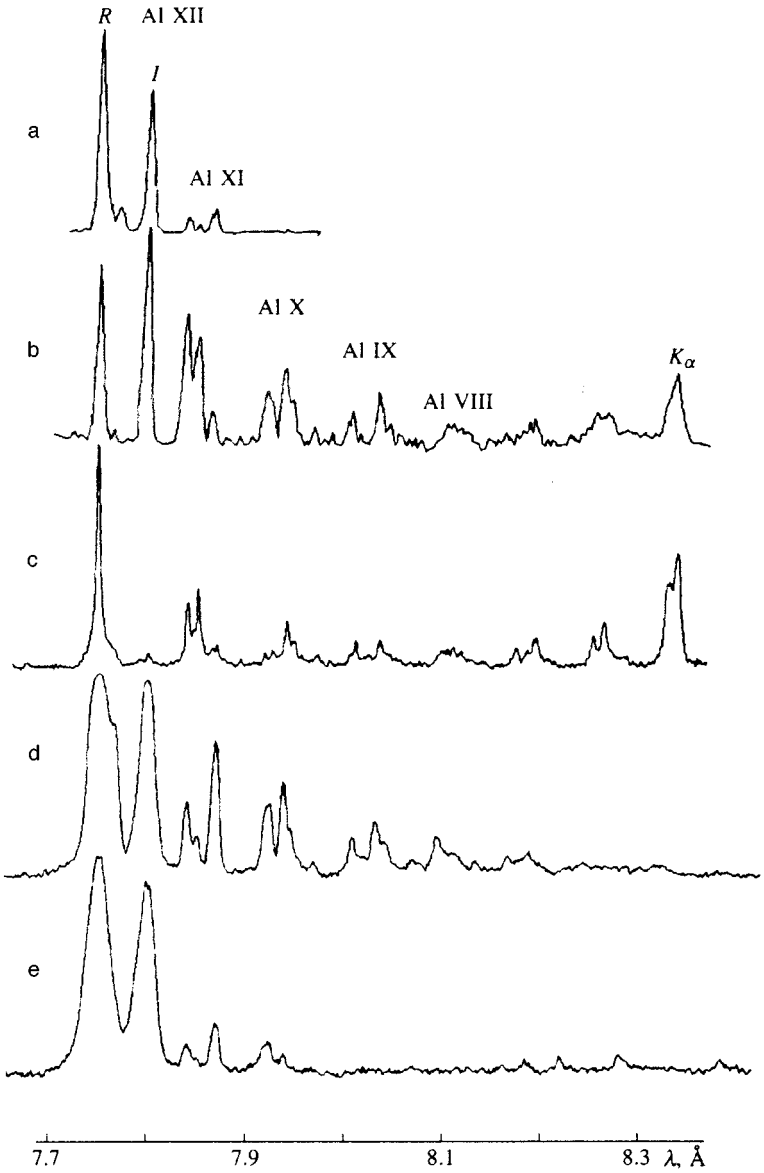

FIG. 1. Density plots of Al plasma, recorded on different machines: a—Nd laser, $\tau=2$ ns, $P=10^{14}$ W/cm^2, $N_e=10^{21}$ cm^{-3}; b—CO_2 laser, $\tau<2$ ns, $P=10^{12}$ W/cm^2, $N_e=10^{19}$ cm^{-3}; c—dye laser, $\tau=80$ fs, $P=10^{17}$ W/cm^2, $N_e=10^{22}$ cm^{-3}; d,e—X-pinch in the presence (d) and absence (e) of the electron beam, $N_e=10^{18}$ cm^{-3} (R is the resonance line, I is the inter-Raman line).

of measuring the parameters of an electron beam in a laser plasma do not exist.

It has been possible to make substantial advances in the solution of this problem in experiments with a plasma of a completely different type, specifically an X-pinch plasma, the X pinch being a variant of the Z pinch, which uses an explosion in a vacuum high-current diode of two or more crossed wires. Because of the specific geometry of the X pinch, it has been possible to distinguish the effect of different factors in it and to directly observe the change in the nature of the spectrum under the action of an electron beam. Figures 1d and 1e show density plots of spectra from a plasma with the same parameters, obtained at the same time in one experiment in the presence and absence of an electron beam. Details of the experiments and their interpretation are presented below.

The present work presents the first experimental results of a study of an electron beam generated in an X pinch, and also experimental and theoretical studies of the effect of an electron beam on the structure of the K spectrum of multiply charged ions.

2. EXPERIMENTAL STUDY OF THE ELECTRON BEAM

A hot dense plasma was created in the XP generator with a current of 470 kA in a pulse of 100-ns duration (Cornell University, USA) and in the BIN generator with a current of 300 kA in a 100-ns pulse (P. N. Lebedev Physical Institute). The plasma is formed in the X pinch during the explosion of crossed wires in a high-voltage diode. The structure of the X-pinch plasma is schematically depicted in Fig. 2. A detailed description of the diagnostic apparatus used to study the X-pinch plasma is given in Ref. 5. Here we note that a high-precision spectroscopic study of the satellite lines became possible with the use of a spherical mica crystal high-current spectrograph possessing a unique combination of

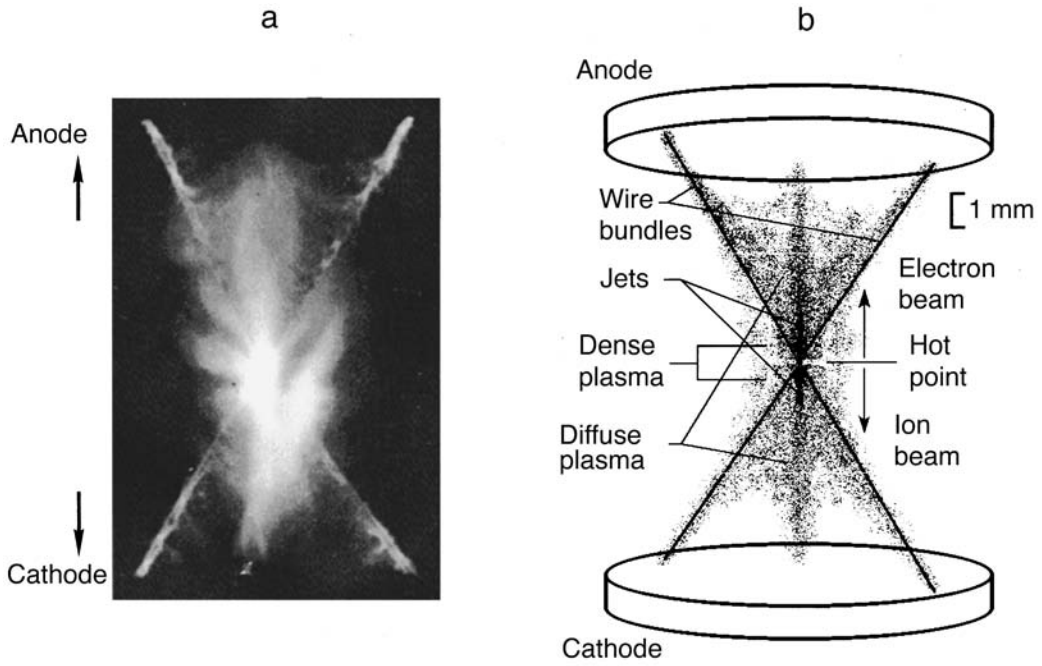

FIG. 2. Pinhole photos of an X-pinch Al plasma (a) and schematic depiction of the X-pinch with the structure of the different plasma regions shown (b) (diameter of the wires $d=25$ μm), obtained through an opening of diameter $D=100$ μm for filter cutoff energy $E>1.2$ keV.

spectral and spatial resolution. The spectral resolution ($\lambda/\Delta\lambda$) can reach 10 000 for a spatial resolution better than 20 μm (Ref. 6).

Studies in recent years have revealed the complexity of the structure of the crossover region of the X pinch and of the processes taking place in it.[7] In this region electron tem-

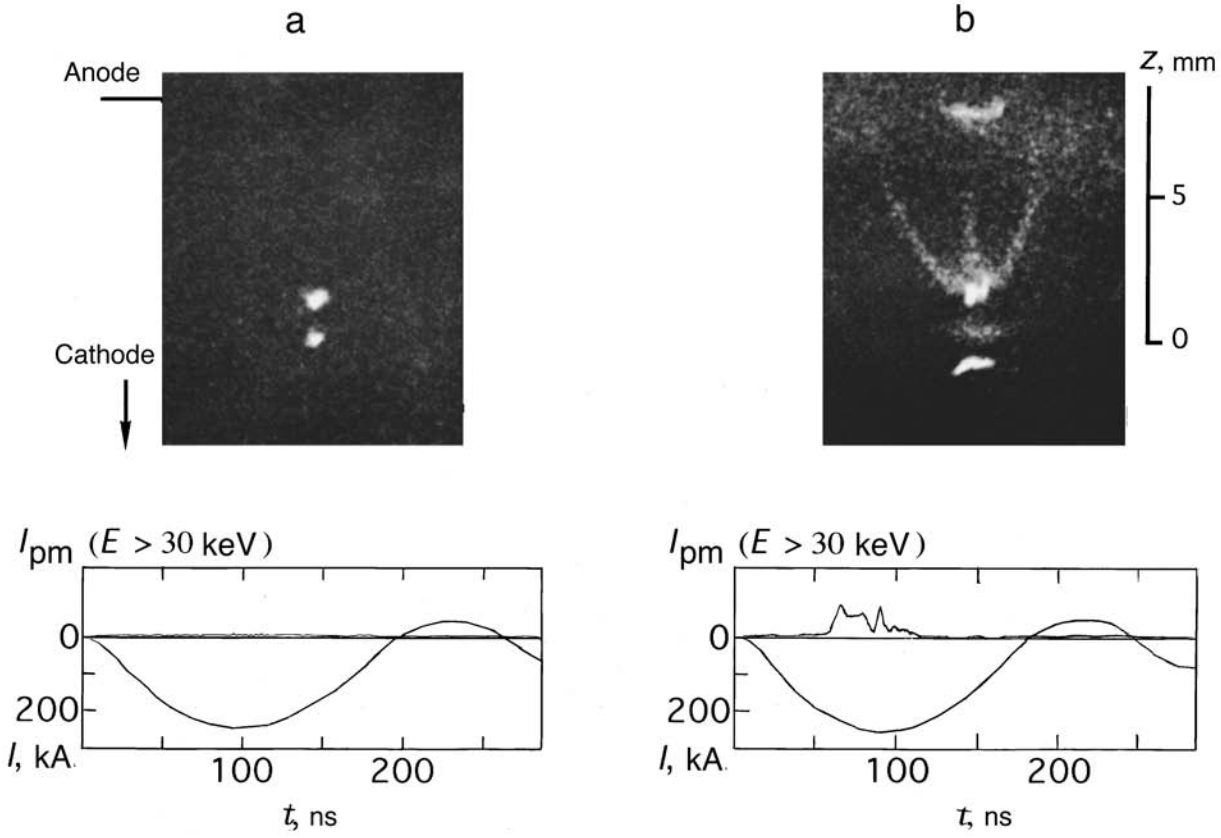

FIG. 3. Pinhole photos of an X-pinch plasma formed by explosion of aluminum wires ($d=35$ μm), $D=80$ μm, $E>8$ keV) and oscillograms of the output current (the lower signal) and the scintillation detector signal: a—without the beam, b—with it.

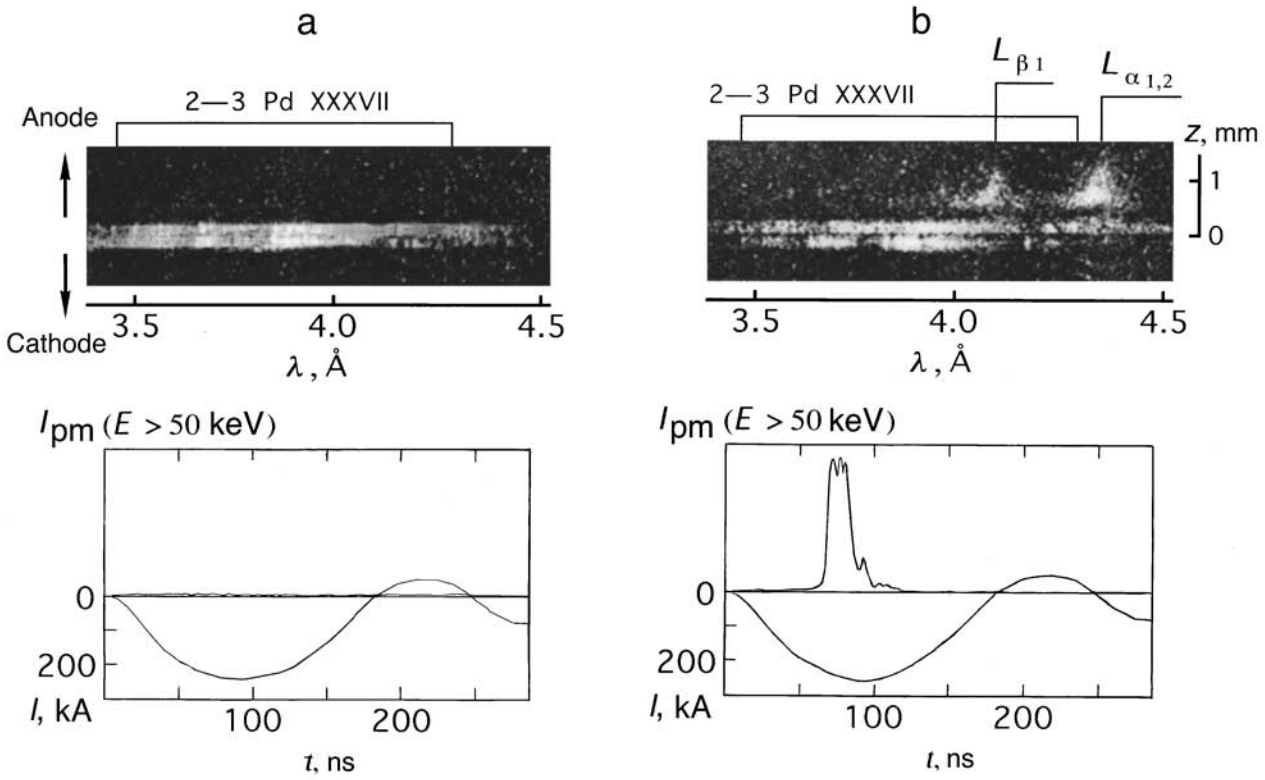

FIG. 4. a—Spectrograms of the Pd L spectrum ($d=20$ μm) recorded with a spectrograph with a convex crystal of CsAP of radius 250 mm, and oscillograms of the output current (lower signal) and the scintillation detector signal without the beam. b—Same as in Fig. 4a, but with the beam.

peratures and densities equal to $T_e \sim 1-2$ keV and $N_e \sim 10^{24}$ cm^{-3} respectively can be observed, and the total emitted energy in the wavelength interval below 100 Å can reach 1 kJ.[8,9] At some distance from the hot point the electron temperature and density fall to $T_e < 100$ eV and $N_e \sim 10^{18}$ cm^{-3}.[8,10]

One peculiarity of the X pinch is the formation during the plasma pinch of the so-called minidiode—a gap in the crossover of the wires—with generation in this spot of a high-energy electron beam (Fig. 2). Estimates give a voltage on the minidiodes of up to 100 kV depending on the conditions of the experiment. Estimates arrived at with the help of pinhole cameras with fixed filters give a maximum energy of $E \sim 15-20$ keV for the electrons generated in the minidiode of an X pinch of light elements ($Z=12-14$) and $E \sim 100$ keV for an X pinch of heavy elements ($Z>29$).[11]

Because of the geometry of the diode and the presence of electromagnetic-field noise in the vacuum diode, it is in practice impossible to carry out direct, quantitative measurements of the beam current at the present time. Conclusions about the presence of an electron beam and estimates of its intensity are usually made on the basis of secondary effects: the emission of the anode and wires from the anode side of the diode (Fig. 3b), higher darkening of films due to beam bremsstrahlung, intense emission of K- and L-series lines of weakly ionized ions (Fig. 4b).

Most accessible for measurement at the present time is bremsstrahlung of the beam electrons. In the experiments on the BIN machine bremsstrahlung of the beam electrons was recorded by means of a scintillation detector with a photomultiplier with a temporal resolution of 8 ns. The detector was positioned outside the vacuum chamber of the setup at a distance of 250 mm from the anode behind filters with different energy cutoffs. We were unable to calibrate the detector and therefore could not carry out exact quantitative measurements; therefore, the measurement data give information only about the relative intensity and hardness of the electron beam and quite rough data about the generation time. Figure 3 displays pinhole photos of an aluminum X-pinch plasma with the beam (Fig. 3b) and without it (Fig. 3a), and also oscillograms of the output current and scintillation detector signal corresponding to these photos. Figure 4 presents spectrograms of the L spectrum of Pd with the beam (Fig. 4b) and without it (Fig. 3a) and the corresponding oscillograms of the output current and scintillation detector signal. From these figures it is clear that the beam intensity and electron energy in the beam are significantly higher in the case of the X pinch arising from the explosion of heavy-element wires and the electron beam exists for 30–50 ns.

Experimental studies have shown that the electron beam strongly influences the nature of the multiply charged ion radiation, especially in the case of light elements. For heavy elements (Cu and heavier) the multiply charged ion spectra and the weakly ionized spectra are spatially separated since the conditions for emission of the multiply charged ion lines are fulfilled only at the hot point, which has dimensions of 10–30 μm, whereas the characteristic K- and L-series lines are usually emitted on the periphery of the diode at a distance of fractions of a millimeter to several millimeters from the hot point (see Fig. 4b) where the beam interacts with the

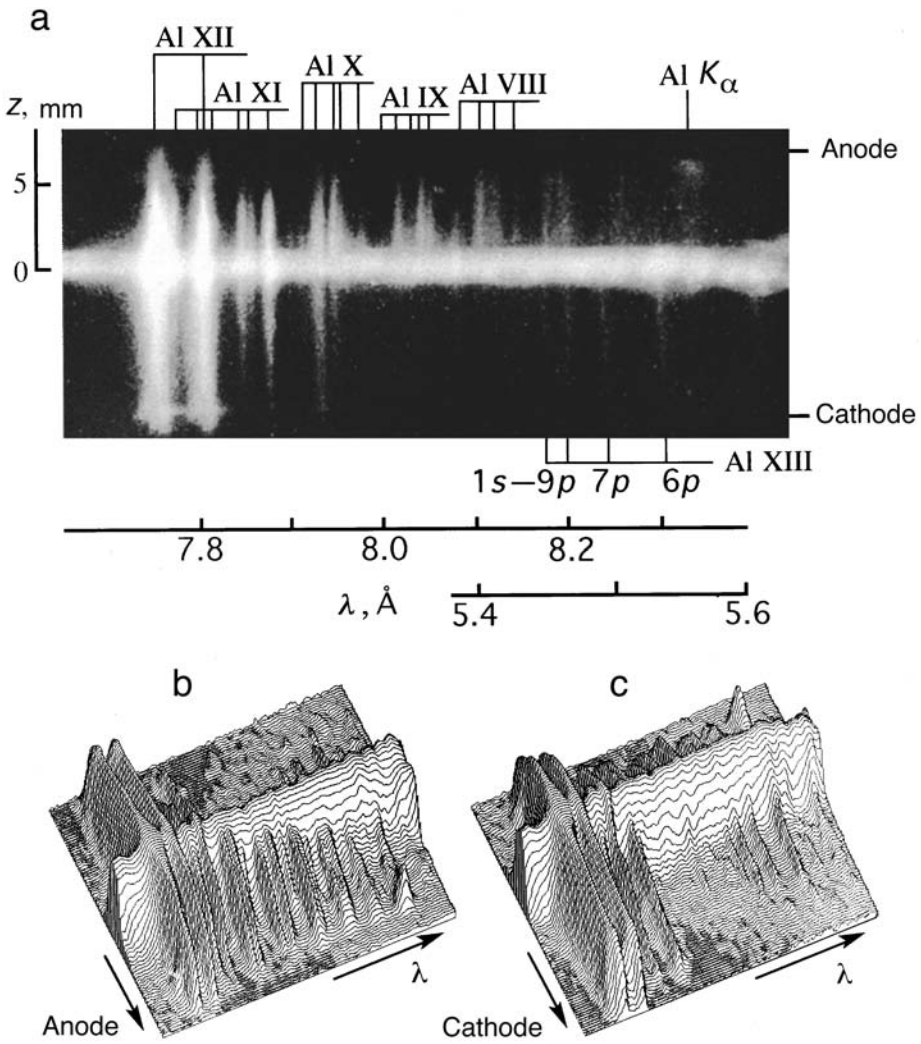

FIG. 5. Spectrograms of radiation from an *X*-pinch plasma from aluminum wires, recorded with a spectrograph with a spherical mica crystal of radius 100 mm on the XP machine (a) and three-dimensional density plots of the spectrum made on the anode (b) and cathode (c) sides of the spectrum.

weakly ionized material. Therefore the effect of an electron beam on the structure of the spectrum of the multiply charged heavy ions is probably small and so far has not been studied. In the case of light elements (Mg, Al, Si) the multiply charged ion lines are observed in an exploding plasma over almost the entire diode (Fig. 5a), but because of the action of the electron beam are significantly more intense from the anode side. Figure 5a displays a time-integrated spectrogram of the emission of an aluminum plasma recorded with the help of a focusing spectrograph with spatial resolution[6] on the XP machine. The resonance line and the inter-Raman line of He-like Al were observed from both sides of the hot point in the second reflection order of the mica crystal. In the anode direction intense satellite lines were also observed corresponding to transitions from Li- to C-like Al ions, whereas from the cathode side these lines are absent or of low intensity. Note that there are a number of elements with intermediate atomic weight (from Ti to Ni), in whose *K* spectra explicit asymmetry was recorded in a number of cases on the anode side in the presence of an intense electron beam (Fig. 6). This is evidence, the same as for the light ions, of the effect of the electron beam, which in this case has a significantly higher intensity and electron energy.

In the case of an *X* pinch of light elements the emission lines of weakly ionized material (Fig. 5a) have low intensity, probably because the beam current and electron energy are significantly smaller (see Figs. 3 and 4) than for an *X* pinch of heavy elements. For example, for an aluminum *X* pinch the *K*-series emission lines of weakly ionized ions have been recorded only very recently using a rapid, focusing spectrograph.

Thus, the specific geometry of the *X* pinch allowed us to isolate the electron beam and study its effect on the radiation of the multiply charged ions and of the weakly ionized material.

3. CALCULATED RESULTS

The unique possibilities of the new spectral apparatus in conjunction with the powerful computing facilities of Los Alamos National Laboratory have made it possible to study the effect of an electron beam on the structure of the *K* spectrum of multiply charged light ions.

With the help of the calculational model described in Refs. 12–15, we perform here a spectroscopic study of the satellite lines observed in an *X*-pinch plasma formed from Al wires (Fig. 5). The population densities of the atomic levels were determined by nonequilibrium quasistationary kinetic

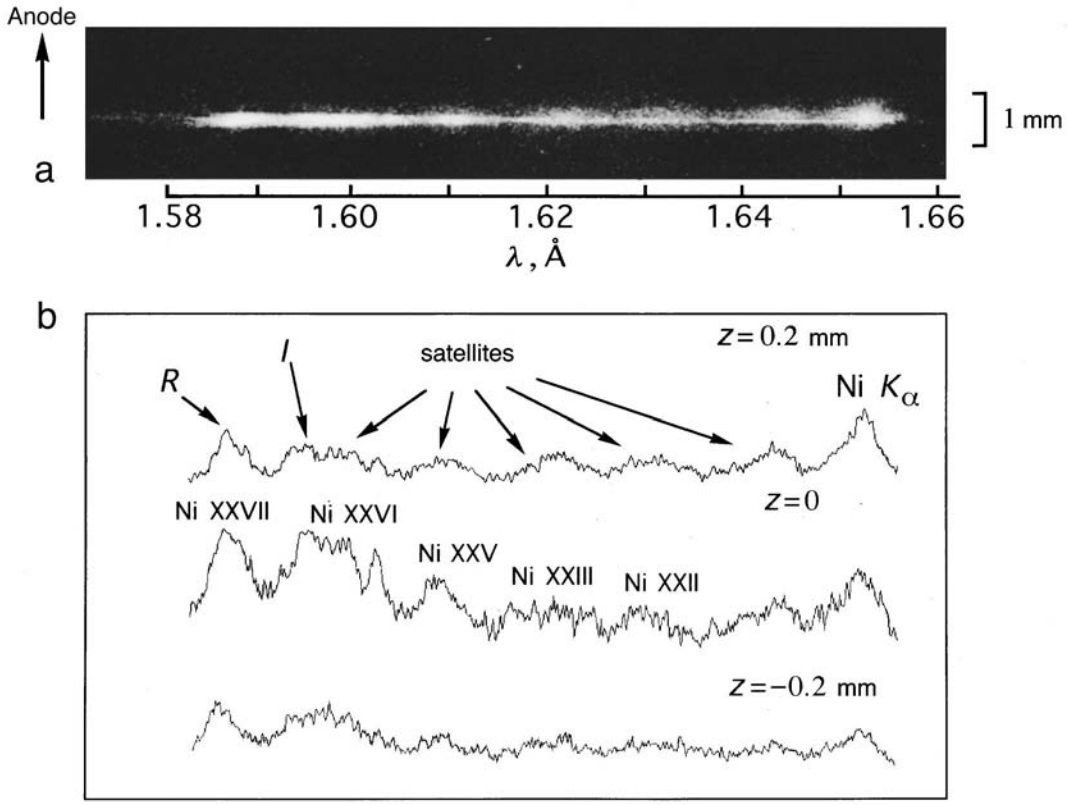

FIG. 6. Spectrograms of the radiation from an X-pinch plasma from nickel wires, recorded with a spectrograph with a spherical mica crystal of radius 100 mm on the XP machine (a) and density plots of the spectrum made in the region of the hot point and at a distance of 0.2 mm from it on the anode and cathode sides of the spectrum (b).

calculations and then used to model the emission spectra. The level structures corresponding to the configurations $1s^2(2s2p)^w$, $1s^2(2s2p)^{(w-1)}3l$, $1s^2(2s2p)^{(w+1)}$, and $1s^1(2s2p)^w3l$ were calculated for Li-, Be-, B-, and C-like ions. Here w is the number of electrons of the shell in the ground state and $(2s2p)^w$ are all the possible permutations of the w electrons on the $2s$ and $2p$ levels. For example, carbon has $w=4$ and $(2s2p)^4$ denotes $2s^22p^2$, $2s^12p^3$, and $2p^4$. The symbol l denotes all possible values of the orbital angular momentum associated with the principal quantum number. Configurations of the type $1s^2$, $1s2l$, $1s3l$, $2l2l'$, and $2l3l'$ were used for the He-like ions. The chosen configurations give an accurate representation of many lines possible in the given spectral range. In addition, the program calculated all the necessary atomic constants. The calculations took into account around 5×10^5 transitions. The electron beam was prescribed by adding a narrow Gaussian distribution, as shown in Fig. 7, to the Maxwellian part of the electron energy distribution function. The calculated spectrum (Fig. 8b) was compared with the experimental density plot corresponding to a distance of 1 mm on the anode side from the crossover (Fig. 8a).

The results of the theoretical calculations depend, generally speaking, on the electron temperature T_e, average beam energy E_0, beam width Γ, the weighting factor f equal to the ratio of the electron density in the beam to the total density

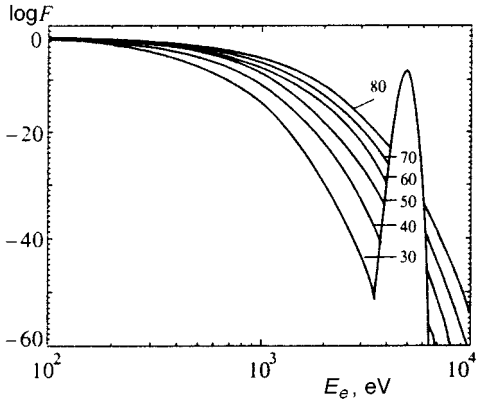

FIG. 7. Electron energy distribution function at various electron temperatures (indicated in kiloelectron volts) for $f=10^{-7}$.

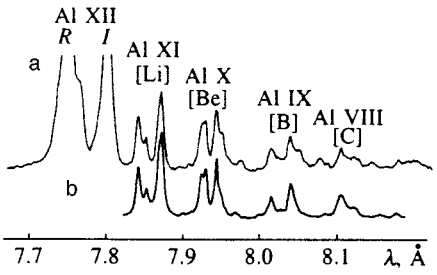

FIG. 8. Comparison of the experimental spectrum (a is the density plot of the spectrum shown in Fig. 6, corresponding to a distance of 1 mm from the hot point on the anode side) with the calculated spectrum (b) in the spectral interval corresponding to the satellite structure of C-...Li-like Al ions.

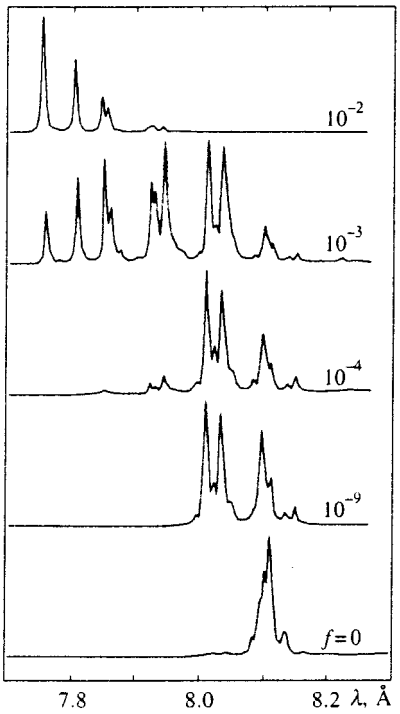

FIG. 9. Calculated emission spectra for an electron temperature of 30 eV and different values of f.

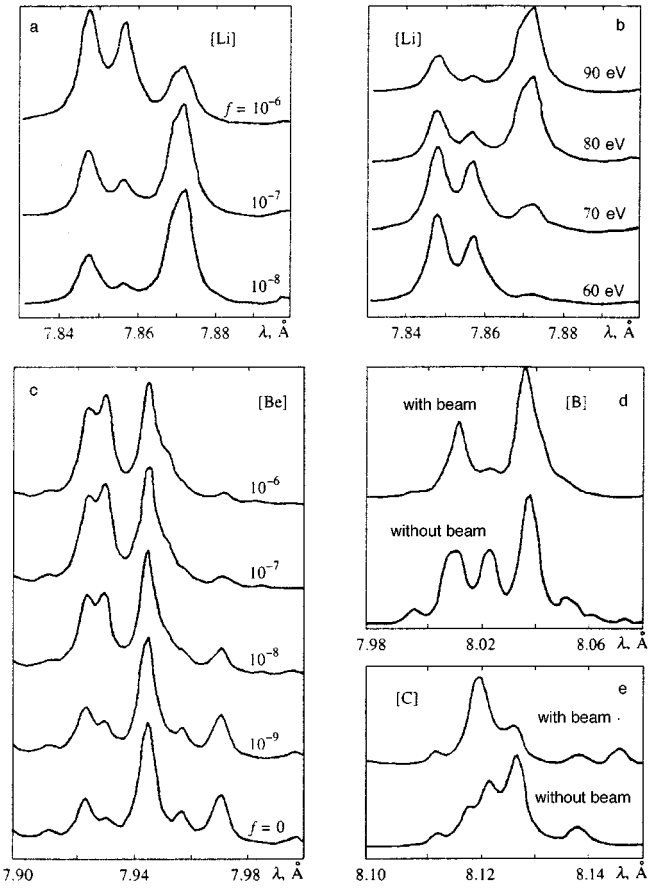

FIG. 10. Calculated emission spectra corresponding to Li-, Be-, B-, and C-like Al ions.

of the electrons included in the calculation, and the electron density N_e. However, the model calculations showed almost no variation in the calculated spectrum in the interval of beam energies E_0 from 5 to 15 keV typical of electron beams generated in an X pinch from light-element wires. This is due to the weak variation of the excitation cross section in this energy interval. Therefore in all subsequent calculations we took $E_e=5$ keV, $\Gamma=100$ eV, and $N_e=10^{18}$ cm^{-3}, which corresponds approximately to the electron density 1 mm from the hot point. Different satellite structures were investigated by varying the electron temperature kT and the weighting factor f.

Figure 9 shows the effect of the electron beam on the radiation of a cold plasma for an electron temperature of 30 eV. The bottom curve is the spectrum in the absence of the beam ($f=0$). In the given case only C-like satellites are present. The spectrum changes substantially even for an insignificant increase in f: for $f=10^{-9}$ B-like satellites appear. In the interval $10^{-9}<f<10^{-4}$ the intensity of the spectral lines does not change significantly, but the total intensity grows even though this cannot be seen in Fig. 9 since all the curves are normalized. For $f>10^{-4}$ the beam becomes sufficiently intense that the ionization balance is shifted, resulting in the appearance of He-, Li-, B-, and Be-like ions. For $f=10^{-2}$ only He- and Li-like ions are formed.

Figure 10 displays calculated satellite structures at different stages of ionization (from C- to Li-like ions). The calculations can be used for a quantitative estimate of the effect of an electron beam on the spectral characteristics of the lines.

Figures 10a and 10b show how the satellite structure of the calculated spectrum for Li-like ions depends on the electron beam intensity and electron temperature, respectively. It can be seen from Fig. 8 that the calculated spectrum corresponding to an electron temperature of $T_e=80$ eV and $f=3\times10^{-7}$ agree best of all with experiment. A small discrepancy between the calculated and experimental wavelengths can be seen. It is customary to denote the lines grouped in the wavelength regions 7.85, 7.86, and 7.88 Å as qr, $abcd$, and jkl. It is well known that the jkl-group lines are populated mainly by dielectronic recombination, whereas the qr- and $abcd$-group lines are more sensitive to excitation by the electron beam of the inner shells of the Li-like ions.

Figure 10c shows how the satellite structure of the calculated spectrum for Be-like ions depends on the electron beam intensity for the electron temperature equal to 60 eV. The calculated spectrum for $f=10^{-7}$ is found to be in beautiful agreement with the experimental spectrum (see Fig. 8). It is clear from the figure that the group of lines in the vicinity of 7.92 Å is pumped by electron collisions to a greater degree than the group of lines around 7.94 Å. The case $f=0$ completely disagrees with experiment.

Figure 10d shows B-like satellites calculated with and without the beam for $T_e=40$ eV and $f=10^{-7}$. It can be seen from comparing with Fig. 8 that the main features of the experimental spectrum are in good agreement with calculation. Note that the central satellite line (8.02 Å) has almost

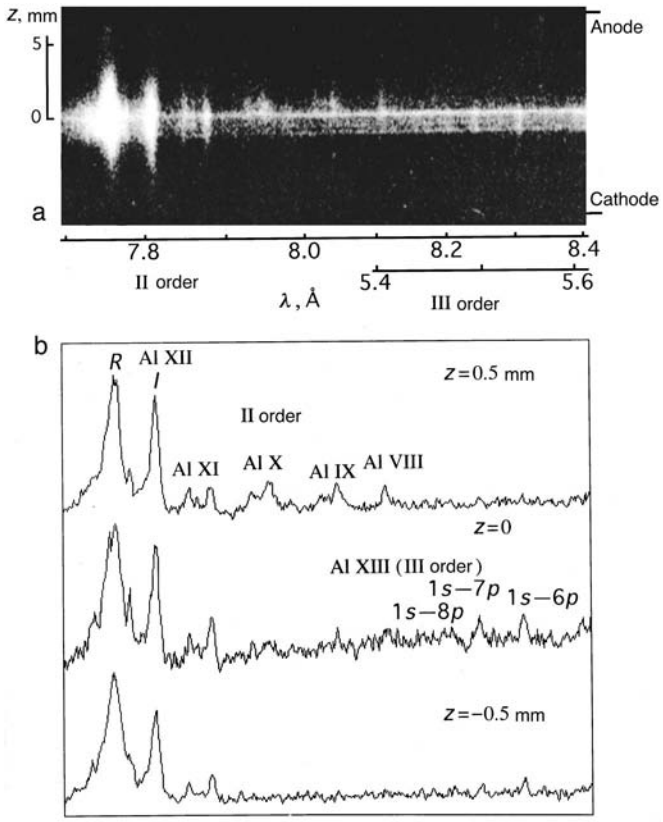

FIG. 11. Spectrogram of the radiation from an X-pinch plasma from aluminum wires, recorded with the help of a detector with a spherical mica crystal of radius 100 mm on the BIN machine (a), and density plots of the spectrum, made in the region of the hot point at a distance of 0.5 mm from the hot point on the anode side and on the cathode side of the spectrum (b).

disappeared while the 8.01 and 8.04 Å lines are pumped by electron collisions.

Figure 10e compares the calculated satellite structures with and without the beam for C-like Al ions. The calculation was carried out for $T_e=30$ eV and $f=10^{-7}$. Note that the given spectra differ strongly and that the calculation with the beam shows significantly better agreement with experiment (see Fig. 8), especially as to the peak intensity of the lines and the structure of the beam profile. For the case without the beam, the lines emitted as a result of dielectronic recombination predominate, and for the case with the beam the lines emitted due to collisional excitation predominate. The lines having the greatest intensity in the presence of the electron beam are almost invisible in its absence.

Note that the strong dependence of the structure of the spectrum on the electron beam support the idea of using it as a diagnostic of beam density.

Figure 11a displays the time-integrated spectrum of the He-like Al ions recorded on the BIN machine using the spectrograph with a spherically curved mica crystal of 100-mm radius. Indirect data, enumerated above, show that the electron beam intensity on the BIN machine is less than on the XP machine; however, the light-ion spectra also have a pronounced asymmetric intensity distribution of the satellite lines in the anode–cathode direction. Figure 11b presents density plots of the spectrum, made at a distance of 0.5 mm above the hot point in the direction of the anode, in the region of the hot point, and 0.5 mm below it in the direction of the cathode. Comparison of the intensities of the satellite lines in the anode direction with the calculated spectra according to the above technique shows that the calculated spectrum agrees well with the experimental spectrum when the beam intensity makes up 10^{-8} to 10^{-7} of the total electron density in the plasma, which is in line with the assumption of a lower beam density on the BIN machine. Note that we have considered time-integrated spectra; therefore the different groups of satellite lines can emit at different times for a different beam intensity and electron temperature.

4. CONCLUSION

In this paper we have presented the first experimental and theoretical studies of the effect of an electron beam generated in an X-pinch plasma on the structure of the multiply charged ion spectrum. In the instance of the aluminum K spectrum, recorded in an X-pinch aluminum plasma on the XP machine at Cornell University, we have succeeded for the first time in theoretically estimating the effect of the electron beam on the structure of the K spectrum of multiply charged, light ions. Calculations of the observed satellite lines of Li-, Be-, B-, and C-like Al ions based on the collisional–radiative model based on a non-Maxwellian electron distribution are found to be in good agreement with the experimental data for the electron beam intensity constituting 10^{-7} of the total electron intensity in the plasma. On the basis of the calculations in this study we also estimated the electron beam intensity in an X pinch of light-element wires on the BIN machine. We conclude that it is possible to compare the calculated and experimental spectra, with the aim of allowing for the effect of the electron beam on the structure of the light-ion spectra over a wide interval of electron beam intensities.

[1] V. A. Boiko, V. V. Vinogradov, A. Ya. Faenov *et al.*, J. Sov. Laser Res. **6(2)**, 85 (1985).
[2] V. A. Boiko, V. V. Vinogradov, S. A. Pikuz *et al.*, J. Phys. B **12**, 213 (1979).
[3] A. Ya. Faenov, S. A. Pikuz, and A. S. Shlyaptseva, Phys. Scr. **49**, 41 (1994).
[4] J. P. Geindre, P. Audebert, A. Rousse *et al.*, Phys. Scr. **53**, 645 (1996).
[5] I. Yu. Skobelev, A. Ya. Faenov, S. A. Pikuz *et al.*, Zh. Eksp. Teor. Fiz. **108**, 1263 (1995) [JETP **81**, 692 (1995)].
[6] A. Ya. Faenov, A. R. Mingaleev, S. A. Pikuz *et al.*, Kvantovaya Elektron. **20**, 457 (1993) [Quantum Electron. **23**, 394 (1993)].
[7] G. V. Ivanenkov, A. R. Mingaleev, S. A. Pikuz *et al.*, Fiz. Plazmy **22**, 403 (1996) [Plasma Phys. Rep. **22**, 363 (1996)].
[8] S. A. Pikuz, B. A. Bryunetkin, G. V. Ivanenkov *et al.*, Kvantovaya Elektron. **20**, 237 (1993) [Quantum Electron. **23**, 201 (1993)].
[9] D. H. Kalantar, D. A. Hammer, K. S. Mittal *et al.*, J. Vac. Sci. Technol. **9**, 3245 (1991).
[10] S. A. Pikuz, V. A. Bryuinetkin, G. V. Ivanenkov *et al.*, J. Quant. Spectrosc. Radiat. Transf. **51**, 291 (1994).
[11] S. A. Pikuz, V. M. Romanova, T. A. Shelkovenko *et al.*, Phys. Scr. **51**, 517 (1995).
[12] J. Abdallah, Jr., A. Ya. Faenov, D. A. Hammer *et al.*, Phys. Scr. **53**, 705 (1996).
[13] F. B. Rosmey, J. Quant. Spectrosc. Radiat. Transf. **51**, 319 (1994).
[14] J. Abdallah, Jr. and R. E. H. Clark, Los Alamos National Laboratory Reports, LA-11436-M, I–V (1988).
[15] J. Abdallah, Jr., R. E. H. Clark, J. M. Peek, and C. J. J. Fontes, J. Quant. Spectrosc. Radiat. Transf. **51**, 1 (1994).

Translated by Paul F. Schippnick

Current–voltage characteristics of extended Josephson junctions with viscous magnetic-flux transport

Yu. E. Kuzovlev

A. A. Galkin Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, 340114 Donetsk, Ukraine

(Submitted 21 December 1996)
Zh. Éksp. Teor. Fiz. **112**, 910–925 (September 1997)

A theory that describes the current–voltage characteristics of extended Josephson junctions is built. The width of the junctions is much larger than the mean free path of a single fluxon, with the length of the path determined by energy dissipation in the junction. Explicit analytic expressions are derived for the maximum supercurrent in the junction and the low-voltage asymptotic behavior of the current–voltage characteristics. © *1997 American Institute of Physics.* [S1063-7761(97)01109-8]

1. INTRODUCTION

Research of phenomena that occur in extended Josephson junctions is extensive. Most of this research, however, deals with junctions with an overlap geometry of the superconducting arms. Here the external transport current in the junction enters into the equation for the jump in the phase of the order parameters between the junction arms in the form of a source evenly distributed along the length. Less studied is the in-line geometry of the arms. Under certain conditions (see Ref. 1) this geometry leads to a model equation of the sine-Gordon type with friction, and no explicit source of current exists. Instead, the transport current is introduced by boundary conditions imposed on the magnetic field at the edges of the junction.

As shown in Ref. 2, in such junctions the resistance mode is characterized by a strong spatial inhomogeneity in the distribution of fluxons (Josephson vortices) if the junction width L is much larger than, first, the Josephson penetration depth λ_j (i.e., the size of a single standing fluxon) and, second, the mean free path λ_f that a fluxon follows before stopping because of dissipation in the junction. The inequality $L \gg \lambda_f$ means that the fluxon motion is of a viscous, nonballistic, nature and that low-amplitude excitations in the junction are strongly damped.

These circumstances lead[2] to a nearly complete disappearance of jumps on the current–voltage characteristic of the junction (Fusko steps), jumps that are related to resonances in fluxon motion and the generation of low-amplitude excitations (''plasma waves'') and are well-pronounced when $L < \lambda_f$. The remaining distinctive features of the current–voltage characteristic are, first, the magnitude of the excess supercurrent in the ohmic region and, second, the shape of the current–voltage characteristic in the low-voltage region, where the current I is lower than it critical value I_c. Here I_c is the maximum current that can be sent through the junction without resistance. The resistance mode may also exist when $I < I_c$, on the so-called single-particle branch of the current–voltage characteristic, which we study here (in practice, this resistance mode is usually observed in a reversed bypass of the current–voltage characteristic).

In this paper we propose an approximate analytic theory of these parameters of the current–voltage characteristics with $L \gg \lambda_f$ and $L \gg \lambda_j$ in the absence of an external magnetic field. First, we show that the excess supercurrent on the single-particle branch has a mildly sloping maximum, and that it tends to zero at low voltages (being almost independent of the junction's width) and at high voltages across the junction (the larger the ratio L/λ_j, the slower the supercurrent approaches zero). Here the behavior of the current–voltage characteristic in the nearest ohmic region is close to that of a linear dependence, which intercepts a line segment on the I axis (the effective excess current) of a length of approximately $2I_c/3$. Second, at low voltages across the junction, when the normal current is much lower than the supercurrent, the current decreases approximately as the reciprocal logarithm of the voltage U, so that the single-particle branch of the current–voltage characteristic has a logarithmic ''beak'' with an infinite derivative at $U=0$. However, this feature of the current–voltage characteristic caused by nonballistic transfer of magnetic flux may become less-pronounces in the junction between two thin (in the direction of the magnetic field) arms because of the nonlocal magnetic interaction of fluxons through ''normal'' space. We also show that the observed[2] ''fine structure'' of time-averaged distributions of field and current is the result of the nonlinear interference between chains of fluxons traveling from the edges to the center of the junction and the plasma waves generated at the edges.

2. THE CONVERGING-WAVE APPROXIMATION

2.1. Dimensionless variables

A derivation of the equation for a Josephson junction (of the superconductor-insulator-superconductor type) between massive arms can be found, for instance, in Ref. 7:

$$\frac{c}{4\pi}H_x = \sigma U_t + gU + j_c \sin\varphi,$$

$$2eU = \hbar\varphi_t, \quad \varphi_x \simeq \frac{2\pi dH}{\Phi_0}.$$

The first equation relates the sum of the capacitive, normal, and Josephson components of the current to the curl of the

magnetic field in the junction's plane. Here H is the field, U is the voltage across the junction, φ is the jump in phase, σ and g are the capacitance and conductance per unit junction area, x is the distance from the middle of the junction, the subscripts x and t indicate differentiation with respect to the coordinate along the junction and with respect to time, d is the effective magnetic thickness of the junction ($d>2\lambda$, with λ the London depth of penetration of the arms by the field), and the quantities c, e, j_c, and Φ_0 have standard meaning. The last formula, which follows from magnetic flux quantization in the superconducting arms and which provides a simple local relationship between the phase gradient and the field, is approximate. It is valid if the characteristic scale λ_j of magnetic field variations in the junction, $\lambda_j \equiv \sqrt{\Phi_0 c/8\pi^2 d j_c}$, is much larger than λ, and this can easily be verified. The given expression for λ_j immediately follows from the above formulas.[1] The integral on the right side of the first equation over the entire junction area is equal to the total transport current I, which makes it possible to express I in terms of the field strength at the junction edges:

$$I = D\frac{c}{4\pi}\left[H\left(\frac{L}{2}\right) - H\left(-\frac{L}{2}\right)\right],$$

where D is the thickness of the junction arms.

To simplify the mathematics, we introduce dimensionless notation. We choose λ_j as the unit of length, and write the third equation and, with its help, the boundary condition imposed on the field in the form

$$\lambda_j \varphi_x = \frac{H}{H_0}, \quad I = I_0 \lambda_j \left[\varphi_x\left(\frac{L}{2}\right) - \varphi_x\left(-\frac{L}{2}\right)\right],$$

where we have introduced the natural units for field and current: $H_0 \equiv \Phi_0/2\pi d\lambda_j$ and $I_0 \equiv D j_c \lambda_j$. For a time unit we take $\tau \equiv \lambda_j/u$, where u is the velocity of Swihart electromagnetic waves in the junction [from the initial equations we can easily find that $u \equiv c(4\pi d\sigma)^{-1/2}$)]. For the units of voltage and resistance we must then take $U_0 = \hbar/2e\tau$ and $R_0 \equiv U_0/I_0 = 4\pi d u/Dc^2$, respectively.

Reducing all the variables to dimensionless form via the units introduced above (replacing x/λ_j by x, I/I_0 by I, etc.), we obtain at the equation

$$\varphi_{xx} = \varphi_{tt} + \gamma\varphi_t + \sin\varphi, \tag{1}$$

with the boundary condition

$$I = \varphi_x\left(\frac{L}{2}\right) - \varphi_x\left(-\frac{L}{2}\right).$$

If we take into account the symmetry of the current distribution with respect to the junction's center $x=0$, we can write the boundary condition in the form of a pair of conditions:

$$\varphi_x\left(t, x = \pm\frac{L}{2}\right) = \pm\frac{I}{2}. \tag{1a}$$

The dimensionless parameter γ in (1), which determines dissipation, is given by the expression

$$\gamma = \frac{\Phi_0 g u}{2\pi j_c \lambda_j c} = \frac{R_0 \lambda_j}{RL},$$

where R stands for the normal junction resistance: $R = (gDL)^{-1}$.

Let us study the mean free path that a fluxon follows in a wide junction before stopping, assuming that the initial fluxon velocity is comparable to u, i.e., the fluxon's maximum velocity. Simple calculations yield $\lambda_f \sim \lambda_j/\gamma = RL/R_0$. Thus, the condition that fluxon motion is viscous is $R \ll R_0$ or, in dimensionless notation,

$$\gamma L \gg 1. \tag{2}$$

Note that for a complete set of junction characteristics we can take the following six parameters: λ, L, and D (which determine the junction geometry); I_0 and R (which specify the scales of the current–voltage characteristics); and u. Since u is lower than the speed of light by a factor of 10 to 15 and assuming that $d \simeq 2\lambda$, we find that $R_0[\text{ohm}] \sim 10-100(\lambda/D)$. If the condition $L \gg 1$ (or $L \gg \lambda_j$ in dimensional form) is satisfied, the critical junction current I_c can be expressed very simply in terms of I_0. Indeed, a stable static solution of Eq. (1) consists of two ''incomplete'' fluxons (of different polarity) originating at the junction edges. The shape of a singular standing fluxon is described by the formula $\varphi_x = 2/\cosh(x - x_0)$. Combining it with (1a), we find that the maximum value of the transport current attainable in the static mode is $I_c = 4$ or, in dimensional form, $I_c = 4I_0 = 4Dj_c\lambda_j$.

2.2. Statement of the problem

In the stationary resistance mode, the time derivative of the phase $\varphi(t,x)$ has a nonzero time-averaged component. Here the constant component of voltage is

$$\langle U \rangle = U_0 \tau \langle \varphi_t \rangle \equiv U_0 \Omega,$$

where the angle brackets stand for time averaging, and Ω is the dimensionless phase drift rate, which obviously coincides with voltage reduced to dimensionless form. If not stated otherwise, dimensionless variables are used throughout.

On the whole, the pattern of magnetic flux transfer in a wide junction observed in the resistance state[2] is too complicated even for an approximate analytic description, although the basics are simple. Vortices that originate at one edge of the junction and antivortices originating at the opposite edge move toward the center, where they annihilate each other. However, in view of condition (2), inertia alone is not sufficient for the vortices and antivortices to reach the center. What is needed is a gradient in the vortex concentration, which creates a pressure gradient acting against the friction force, i.e., toward the center. Hence the time-periodic (or quasiperiodic) waves of magnetic field, $\varphi_x(t,x)$, and local voltage, $\varphi_t(t,x)$, traveling together with the chain of vortices toward the center are spatially inhomogeneous. Inhomogeneity means that the amplitude and shape of the temporal oscillations of field and voltage depend on the point in the junction.

What complicates matters is the fact that low-amplitude (''plasma'') excitations are superimposed on the traveling chain of vortices. When there are many vortices in the junction, the excitations manifest themselves as ''acoustic'' waves of vortex concentration, which are generated at the

center of the junction in vortex–antivortex annihilation and at the edges. As a result, the vortices that are close to these two sources of ''plasma'' waves can be in retrogressive motion, which cannot be described in terms of a wave travelling from the edges to the center.

Note, however, that computer modeling (see also Sec. 3) shows that plasma wave generation occurs primarily because the boundary conditions (1a) rigidly fix the instantaneous value of the total current, and this must remain unchanged when a vortex enters the junction. The picture is simplified considerably if we replace the conditions (1a) by less stringent, smoothed over time, boundary conditions:

$$\left\langle \varphi_x\left(t, x=\pm\frac{L}{2}\right)\right\rangle = \pm\frac{I}{2}, \tag{3}$$

which specify only the constant component of the total current. More precisely, now everywhere, with the exception of a narrow middle region of width of order unity, retrogressive motion disappears, and what is left is a monotonically converging chain of vortices. Retrogressive motion at the center occurs because an annihilating vortex–antivortex pair performs oscillations in the course of annihilation. Of course, at the center the converging wave pattern is sure to be disrupted, so that this region merits a separate study.

The conditions (3) make it possible to construct a fairly simple approximate solution of Eq. (1) for a real junction of finite width in the form of a fragment, referring to the region $|x|<L/2$, of a wave of vortices that converges from infinity in an infinitely wide junction [in this way we implicitly introduce additional restrictions that replace the more detailed conditions (1a) for the instantaneous current]. Going back to (1a) from (3), we can build a solution in the form of a linear combination of the given approximation and plasma waves, which, by their definition, do not carry a ''topological charge,'' i.e., do not change the number of vortices in the junction (at least, on the average in time).

We can expect that this modification of solution will have a small effect on the current–voltage characteristic, since the constant component of the total current is fixed to the same extent by the conditions (1a) and by the conditions (3), while the number of vortices is fixed neither by (1a) nor by (3). Because of the high viscosity condition (2), the distribution of the magnetic flux, the dissipation in the junction, and hence the current–voltage characteristic are determined by the balance between the friction force and the vortex repulsion in the bulk of the junction and are insensitive to the annihilation and vortex entry. In this case, going back to (1a) from (3) leads only to a redistribution of the current density inside the junction and, in particular, to the fine structure mentioned earlier, without changing the time-average of the integral current, and the features of the current–voltage characteristic we are interested in can be analyzed within the less stringent boundary conditions (3).

On the other hand, the approximation based on (3) has independent physical meaning, since, first, in variations of current–voltage characteristics in the fixed-current mode only the low-frequency component of the current is actually monitored, and, second, the instantaneous current in the external circuit may include a variable bias current which is disregarded in (1) and (1a). Moreover, the conditions (3) are even more natural than (1a) for the fixed-voltage mode. In this connection, we note that a numerical experiment, which obeys the principle of minimum entropy production, in the fixed-current mode leads to the ''superconducting'' branch in the current–voltage characteristic (on which $I\to I_c$ as $U\to 0$), while in the fixed-voltage mode it leads to the single-particle branch (which is lower, in current, than the superconducting branch).

2.3. Mathematical formulation of the approximation

Deep in the resistance region, where the time average of the normal current,

$$I_n=\gamma\int\langle\varphi_t(t,x)\rangle dx=\gamma\Omega L,$$

is much higher than the average supercurrent

$$I_s=\int\langle\sin\,\varphi(t,x)\rangle dx,$$

so that the latter can be ignored, we have an approximate solution of Eq. (1) in the form of a converging wave:

$$\varphi(t,x)\simeq\Omega\left(t+\frac{\gamma x^2}{2}\right).$$

The corresponding time averages of the magnetic field, $h(x)\equiv\langle\varphi_x\rangle$, and of the local vortex concentration, $n(x)$, are linear functions of position:

$$h(x)=2\pi n(x)\simeq\gamma\Omega x.$$

Here the local vortex velocity $v=dX(t)/dt$, with $X(t)$ specifying the position of the vortex, at $X(t)=x$ satisfies the phase-constancy condition $\varphi(t,X(t))=\text{const}$ and is equal to $v(x)=-(\gamma x)^{-1}$. The velocity satisfies the necessary condition of magnetic-flux conservation:

$$2\pi n(x)v(x)=h(x)v(x)=-\Omega=\text{const}, \tag{4}$$

i.e., the independence of the transferred, per unit time on the average, magnetic flux of the place in the junction. Obviously, in view of (2), the vortex velocity is always low at the edges.

In the general case, where the I_n-to-I_s ratio is arbitrary, we also seek the solution of Eq. (1) with conditions (3) in the form of a monotonically (i.e., without change of direction) converging chain of vortices and a symmetric opposed ''antichain.'' The central region must be considered separately because always $h(0)=0$, because of the antisymmetric nature of the field distribution, and the vortex velocity, interpreted from the viewpoint of the ''hydrodynamic'' relationship (4), assumes nonphysical values in this region. For the time being, however, we exclude only the point $x=0$ from the converging wave approximation, assuming that our junction is the inner part of an infinite junction. We fix the voltage $\Omega\equiv\langle\varphi_t\rangle$ and find the constant component of the total current from (3).

Accordingly, we seek $\varphi(t,x)$ in the form of a time-periodic converging wave. Allowing for the fact that in a

stationary resistance state, $\varphi(t,x)$ builds up, on the average, in proportion to time, the phase of the wave should be written as

$$\tau(t,x)=\Omega t+\theta(x), \tag{5}$$

and so the solution of Eq. (1) is sought in the form

$$\varphi(t,x)=\tau(t,x)+\Phi(\tau(t,x),x), \tag{6}$$

where the function $\Phi(\tau,x)$ is periodic in the argument τ with a period of 2π (or an integral multiple of 2π). The meaning of the phase $\tau(t,x)$ is that it simultaneously and completely determines small-scale spatial oscillations and high-frequency temporal oscillations of the instantaneous field φ_x and the voltage φ_t. Hence, after we go over to the new pair of variables $\{\tau,x\}$, the role of the variable x changes: now it describes only the relatively large-scale inhomogeneity of the wave, i.e., the dependence of the parameters of fast oscillations on the point in the junction.

Of course, formally expression (6) is still exact. The main idea of using the converging wave approximation is to introduce the additional requirement that the constant component $\Phi(\tau,x)$ is x-independent:

$$\varphi_0\equiv\langle\Phi(\tau,x)\rangle=\text{const}, \tag{7}$$

where the angle brackets stand for averaging over the period of τ. Indeed, if condition (7) is not met, the two terms on the right-hand side of Eq. (6) actually have two different wave phases (τ in the second term and $\tau+\langle\Phi\rangle$ in the first). From the purely mathematical viewpoint, the condition (7) defines the approximation. In view of (7) we have $\langle\Phi_\tau\rangle=0$ and, in addition, the following condition is satisfied:

$$\langle\Phi_x\rangle=\frac{d}{dx}\langle\Phi\rangle=0, \tag{8}$$

from which follows the relationship between the average magnetic field and the phase (5) of the converging wave:

$$h(x)=\langle\varphi_x\rangle=\langle(1+\Phi_\tau)\theta_x(x)+\Phi_x\rangle=\theta_x(x).$$

The vortex velocity, determined, as before, from the condition that the wave phase is constant, $\tau(t,x)=\text{const}$, is

$$v(x)=-\frac{\Omega}{\theta_x(x)}=-\frac{\Omega}{h(x)}.$$

As these expressions for the velocity and field show, the condition (7) automatically guarantees that the converging wave approximation and the hydrodynamic relationship (4) agree. Moreover, if we employ the relationship [which follows from the average of Eq. (1)]

$$\frac{d}{dx}h(x)=\gamma\Omega+\langle\sin\varphi\rangle\equiv j, \tag{9}$$

where j is the average total current density, and combine it with (7) and (8), from (9) we obtain at a direct relationship between the phase $\theta(x)$ and the current–voltage characteristic:

$$I=2h\left(\frac{L}{2}\right). \tag{10}$$

In terms of the new variables, Eq. (1) acquires the following form:

$$h^2\Phi_{\tau\tau}-\Omega^2\Phi_{\tau\tau}+2h\Phi_{\tau x}+\Phi_{xx}+(h_x-\gamma\Omega)(1+\Phi_\tau)=\sin(\tau+\Phi). \tag{11}$$

Of course, the solution of this formally exact equation does not necessarily meet the condition (7). We will show, however, that far from the center all the terms except the first can be ignored on the left side of Eq. (11), and the solution of the simplified equation

$$h^2\Phi_{\tau\tau}=\sin(\tau+\Phi) \tag{12}$$

meets the key condition (7) of the converging wave approximation. First, we note that the amplitude of the oscillations of $\Phi(\tau,x)$ is a decreasing function of $|x|$. Indeed, at a certain characteristic distance x_0 from the center, where the average field $h(x)$ in its buildup begins to exceed 2π, the vortex density becomes, in accordance with (4), much greater than unity, i.e., the reciprocal size of a single vortex. Hence the amplitude of small-scale inhomogeneities in the field and voltage, which stem from the alternation of vortices, must decrease. Since $h(x)$ increases approximately by a linear law, this decrease must proceed by a power law. Consequently, each differentiation of $\Phi(\tau,x)$ with respect to x gives rise to a factor of order $1/|x|$ (while differentiation with respect to τ yields a factor of order unity), and in (11) the term with the greatest weight is the one with smallest number of differentiations in x and the higher power of h. For this reason, the second, third, and fourth terms are small compared to the first. We can also hope that the fifth term is relatively small, since the difference $h_x-\gamma\Omega=\langle\sin(\tau+\Phi)\rangle$, or the average supercurrent density, in it must be small everywhere except, perhaps, in a small region at the center, since the integral supercurrent cannot exceed several units. Therefore, we are left with the first term (which is of the order of unity, just as the right side of Eq. (11) is), so that we have Eq. (12). Below we will additionally verify the validity of these ideas.

The 2π-periodic solution of Eq. (12), as is known, is

$$\Phi(\tau,x)=\Phi_0(\tau,k)\equiv\pi+2\,\text{am}\left(k,\frac{K(k)\tau}{\pi}\right)-\tau=\pi+2\sum_{n=1}^{\infty}\frac{\sin n\tau}{n\cosh ng},$$

$$g\equiv\frac{\pi K((1-k^2)^{1/2})}{K(k)}; \tag{13}$$

i.e., the constant φ_0 defined earlier is equal to π. Here the function $\text{am}(k,u)$ is the elliptic amplitude, $K(k)$ is the complete elliptic integral of the first kind, and the elliptic modulus $k=k(x)$ is related to the parameter $h=h(x)$ (in view of the independence of the period 2π of the coordinate and h) as follows:

$$h=\frac{\pi}{kK(k)}. \tag{14}$$

Combining this expression and (10), we obtain at the following expression for the current–voltage characteristic:

$$I=\frac{2\pi}{qK(q)},\quad q\equiv k\left(\frac{L}{2}\right). \tag{15}$$

3. THE KINETIC EQUATION AND THE CURRENT–VOLTAGE CHARACTERISTIC

By itself, Eq. (12) says nothing about the dependence of the field h and the modulus k on position. But since the modulus determines the size and shape of the temporal oscillations of field and voltage and, thereby, the energy density averaged over the period, this dependence can be found from the exact energy balance relationship

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\varphi_t^2+\frac{1}{2}\varphi_x^2-\cos\,\varphi\right)=\frac{\partial}{\partial x}\varphi_x\varphi_t-\gamma\varphi_t^2 \tag{16}$$

in its time-averaged form

$$\frac{d}{dx}\langle\varphi_t\varphi_x\rangle=\gamma\langle\varphi_t^2\rangle, \tag{17}$$

which relates the divergence of the energy flux (the Poynting vector) and the dissipative losses of energy. Correspondingly, the current–voltage characteristic (15) reflects the kinetics of transfer and dissipation of the oscillation energy. A convenient property of the converging wave approximation is the fact that in this approximation Eq. (17) leads to a closed kinetic equation for $k(x)$.

In terms of the (τ,x) variables, Eq. (17) becomes

$$\frac{d}{dx}\Omega\langle\varphi_\tau(\theta_x\varphi_\tau+\varphi_x)\rangle=\gamma\Omega^2\langle\varphi_\tau^2\rangle. \tag{18}$$

Substituting $\varphi(t,x)$ in the form determined by Eqs. (6) and (13), we can easily see that in the converging wave approximation the condition

$$\langle\varphi_\tau\varphi_x\rangle=0 \tag{19}$$

is valid for any function $k(x)$. The condition follows from the fact that only the amplitudes, and not the phases, in the Fourier series (13) are functions of the modulus and, hence, of the coordinate. Consequently, Eq. (18) together with the relation $h(x)=\theta_x(x)$ obtained earlier lead to the equation

$$\frac{d}{dx}h\langle\varphi_\tau^2\rangle=\gamma\Omega\langle\varphi_\tau^2\rangle. \tag{20}$$

The standard formulas of the theory of Jacobi's elliptic functions[3] yield

$$\langle\varphi_\tau^2\rangle=1+\left\langle\left(\frac{\partial\Phi_0}{\partial\tau}\right)^2\right\rangle=\left(\frac{2}{\pi}\right)^2E(k)K(k)\equiv f(k), \tag{21}$$

where $E(k)$ is the complete elliptic integral of the second kind. Plugging (14) and (21) into (20) and employing the formula[3]

$$\frac{d[E(k)/k]}{dk}=-\frac{K(k)}{k^2},$$

we obtain at a kinetic equation for the modulus:

$$\frac{dk}{dx}=-\frac{\gamma\Omega k^2E(k)}{\pi}. \tag{22}$$

Note that the nontrivial part of $k(x)$, the function $h(x)$, represents the spatial distribution of the average supercurrent: if we were to ignore the supercurrent, $h(x)$ would be a purely linear function, as (9) clearly implies. In the resistance mode, the supercurrent is closely related to the amplitude of oscillations. For instance, with the boundary conditions (1a) the exact equation (17), after we integrate along the length of the junction and perform simple transformations with allowance for the symmetry of $\varphi(t,x)$ with respect to point $x=0$, yields

$$I-\gamma\Omega L=I_s=\gamma/\Omega^{-1}\int\langle(\varphi_t-\langle\varphi_t\rangle)^2\rangle dx.$$

Thus, the supercurrent is expressed in terms of the variance of fluctuations of the voltage φ_t near its average value Ω or, in other words, in terms of ''excess'' dissipation accompanying voltage fluctuations. But variance is insensitive to the phases of Fourier harmonics. This remark helps us understand why a current–voltage characteristic can be found for the averaged energy relation (20), in which the phases also play no role.

To calculate $k(L/2)$ in (15) we must first specify the initial conditions to Eq. (22) at the center of the junction. First, we examine the current–voltage characteristic by ignoring the fact that in the annihilation region the approximation specified by (12) and (13) becomes invalid and by applying Eq. (22) to the entire half of the junction, $0<x<L/2$. Since $h(0)=0$, according to (14) we can set $k(0)=1$. The corresponding solution of Eq. (22) is a function of only one variable, the product $\gamma\Omega x$, and the current–voltage characteristic (15) is a function of the combination $I_n=\gamma\Omega L$, i.e., a function of the normal current: $I=F(I_n)$, where $F(z)$ is determined implicitly by a pair of relations,

$$F=\frac{2\pi}{qK(q)},\quad \int_q^1\frac{dk}{k^2E(k)}=\frac{z}{2\pi}.$$

In dimensional notation, the current–voltage characteristic assumes the form

$$I=I_0F\left(\frac{U}{RI_0}\right),$$

where the current scale $I_0\sim I_c/4$ was defined earlier. Near the center the elliptic modulus varies as

$$k(x)\cong\frac{1}{1+\gamma\Omega x/\pi}.$$

Using the asymptotic behavior of the elliptic integral,

$$K(k)\simeq\frac{1}{2}\ln\frac{8}{1-k},\quad 1-k\ll 1, \tag{23}$$

for voltages so low that the total normal current I_n is much weaker than 2π, we obtain from Eq. (15) expression

$$I\approx\frac{4\pi}{\ln(16\pi/I_n)},\quad I_n=\gamma\Omega L\ll 2\pi. \tag{24}$$

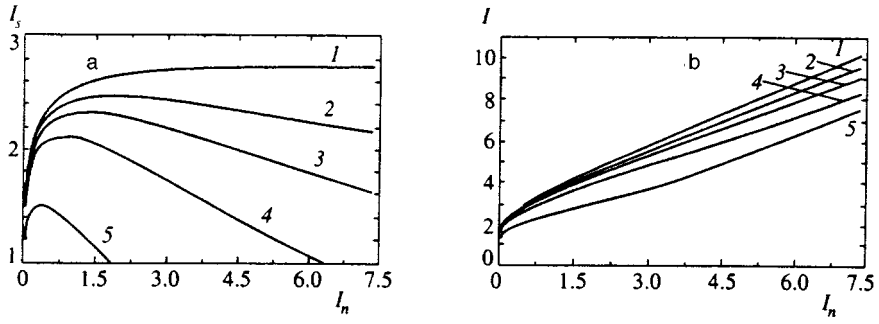

FIG. 1. Dependence of supercurrent (a) and total current (b) on the normal current for different values of the parameter $\delta = x_0/L = 0$ (curves *1*), 0.05 (curves *2*), 0.1 (curves *3*), 0.2 (curves *4*), and 0.5 (curves *5*). To go over to dimensional notation one should replace I by I/I_0, I_s by I_s/I_0, and I_n by U/RI_0, where $I_0 = j_c D\lambda_j$.

Thus, the quasiparticle branch of the current–voltage characteristic has a logarithmic ''beak'' with an infinitely increasing slope near the zero voltage. For the current–voltage characteristic at values of the normal current greater than unity, we obtain from Eqs. (15) and (22) the expression $I \simeq I_n + I_{\rm ex}$, where

$$I_{\rm ex} = 4\left[1 - \int_0^1 \left(\frac{\pi}{2E(u)} - 1\right)\frac{du}{u^2}\right] \approx 2.76 \approx \frac{2I_c}{3} \tag{25}$$

is the asymptotic value of the excess supercurrent.

4. CORRECTION OF THE CONVERGING-WAVE APPROXIMATION

4.1. Contribution of the annihilation region

Before we discuss the current–voltage characteristics with allowance for the special role of the central region, let us estimate the width of this region. At its boundary the discarded terms in Eq. (11) must be of the order of unity. Plugging $k(x)$ in the form of a solution of Eq. (22) with the condition $k(0) = 1$, we obtain at the following estimate for, say, the third term on the left side of Eq. (11):

$$2h\frac{\partial^2\Phi_0}{\partial\tau\partial x} = 2h\frac{\partial^2\Phi_0}{\partial\tau\partial k}\frac{dk}{dx} \sim \frac{h}{1-k^2}\frac{dk}{dx} < \frac{1}{|x|}.$$

For the average supercurrent density $j_s = \langle \sin\varphi \rangle$ at the center, Eqs. (9), (14), and (23) yield

$$\begin{aligned} j &= \frac{dh}{dx} = \frac{\gamma\Omega E^2(k)}{(1-k^2)K^2(k)}, \\ j_s &= j - \gamma\Omega \sim \frac{2\pi}{|x|}\left[\ln\frac{8\pi}{\gamma\Omega x}\right]^{-2}. \end{aligned} \tag{26}$$

We see that in the adopted approximation the third term becomes roughly equal to unity, and j_s acquires nonphysical values greater than unity at approximately $|x| < 1$; i.e., the width x_0 of the region where the approximations (12) and (13) do not work is fairly small, $x_0 \sim 1$, and is weakly dependent on $\gamma\Omega$ (due to the slow growth oscillations at the center).

On the whole, differentiation of elliptic functions with respect to x yields, with allowance for Eq. (22) and the asymptotic behavior (23), a factor of the order of

$$\frac{1}{k}\frac{dk}{dx} \sim \frac{\gamma\Omega k}{2} < \frac{1}{|x|} \quad \text{for} \quad k \ll 1,$$

when $\gamma\Omega|x|/\pi > 1$, i.e., in the region of small fluctuations, and a factor of the order of

$$\frac{1}{1-k}\frac{dk}{dx} \sim \frac{1}{|x|} \quad \text{as} \quad k \to 1,$$

when $\gamma\Omega|x|/\pi < 1$, i.e., in the region where fluctuations are not small. Hence the transition from (11) to (12) is justified everywhere when $|x| > x_0$, where $x_0 \sim 1$, and even when oscillations are strong over the entire width of the junction.

What we have said so far also implies that in the annihilation region the average supercurrent density has a maximum (and may reach values of order unity). Since this region is narrow, we approximate the average total-current density in it by a constant, $j(x) \simeq j_0$, so that $h(x) = j_0 x$, and apply the converging wave approximation for $|x| > x_0$. Here we find the initial condition to Eq. (22), $k(x_0) \equiv k_0$, and the quantity j_0 from the condition that the field and current density are continuous at the points $\pm x_0$. Excluding j_0 from the pair of the corresponding relations via (14) and (26), we can easily derive the following single matching equation for determining k_0:

$$I_n k_0 E^2(k_0)\,\delta = \pi(1-k_0^2)K(k_0). \tag{27}$$

Here we have introduced the dimensionless parameter $\delta \equiv x_0/L$.

Obviously, instead of a single universal current–voltage characteristic $I = F(I_n)$ we have, in terms of the variables (I, I_n), different characteristics for junctions of different lengths. Here the above current–voltage characteristic with the asymptotic behavior (24) and (25) corresponds to the limit where $\delta \to 0$, i.e., an infinitely long junction. The role of the parameter δ becomes clear from Fig. 1, where we plot the current–voltage characteristics, $I(I_n, \delta)$ (Fig. 1b), and the dependence of the supercurrent, $I_s(I_n, \delta) = I - I_n$, as functions of the normal current for different values of δ (Fig. 1a), where at larger values of δ the supercurrent's contribution is smaller and the curves lie lower. The lowest curve for the formal limit of δ emphasizes the tendency of a current–voltage characteristic to change its shape: as the junction becomes shorter and δ grows, the excess supercurrent in the ''ohmic'' region decreases progressively faster with increasing voltage, and the bending of the current–voltage characteristic increases progressively. Clearly, the excess current (25) is the maximum possible supercurrent on the ''resistance'' branch.

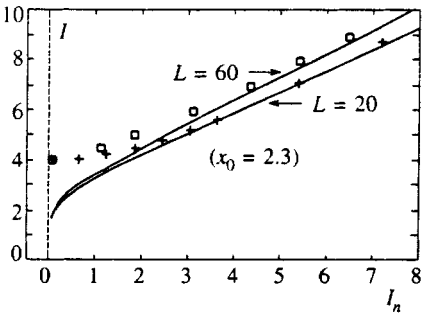

FIG. 2. Two branches of the current–voltage characteristic: the computer-modeled (Ref. 1) upper superconducting branch (+ and □) and the lower, resistance, branch calculated at $x_0=2.3$ by the formulas of the present paper (solid curves). Both branches were constructed for two junction lengths: $L=20$ (+ and lower curve) and $L=60$ (□ and upper curve).

For the correction of the converging wave approximation to acquire a quantitative form, we must find the parameter x_0 it introduces, or the half-width of the middle annihilation region. At this point, in our discussion we known only its order of magnitude. Here we will not attempt to solve the problem analytically. Instead, we will try to establish the value of x_0 from computer modeling data.[2] In view of the above estimates, it is logical to use the approximation in which x_0 is constant (i.e., independent of γ, I_n, and L). In this case we can find it from the requirement that in the near ''ohmic'' region (i.e., for normal currents I_n amounting to several units) two branches of the current–voltage characteristic, the one we consider here and the one studied in Ref. 2 (which tends to I_c for small I_n), merge. If the assumption that x_0 is constant is true, the value of x_0 found in this manner must be the same for different junction lengths. We found that this is exactly the case: the comparison of the data of Ref. 2 and the results of calculations that used Eqs. (15), (22), and (27) for the junction lengths $L=20$, 30, 40, and 60 yield close values of x_0: precisely, $x_0\simeq 2.3$. For example, Fig. 2 shows the lower branch of the current–voltage characteristic calculated at $x_0=2.3$, and the upper branch from Ref. 2 for $L=20$ (+ and the lower curve) and $L=60$ (□ and the upper curve). Figure 2 shows that the excess current, which is defined as the line segment intercepted on the axis $I_n=0$ by the linear continuation of the closest ohmic segment of the current–voltage characteristic, is only slightly smaller than $I_{\rm ex}$, irrespective of the value of δ.

We also note that the parameter δ corresponding to the width $L=60$ is already so small that the corresponding current–voltage characteristic is virtually indistinguishable from the current–voltage characteristic for a very wide junction ($\delta=0$), for which the value of x_0 is unimportant. Hence the fact that in the ohmic region the single-particle branch calculated here and the experimental superconducting branch coincide supports the validity of Eq. (25).

4.2. Role of the finiteness of the thickness of junction arms

We wish to study the behavior of current–voltage characteristics when voltages tend to zero, with the leading contribution provided by the supercurrent. As Eq. (27) and Figs. 1 and 2 show, even with allowance for the correction the characteristic logarithmic ''beak'' is present, irrespective of the values of x_0 and δ. However, as noted in Ref. 4, the derivative of the current–voltage characteristic, as $I_n\to 0$, may become finite because of a spatially nonlocal relationship between the magnetic field and the current density in the junction, a feature not allowed by the model (1) but nevertheless always present. Let us discuss this correction of the current–voltage characteristic in detail.

In a junction, whose superconducting arms have a finite thickness D in the direction of vortex orientation, vortex dynamics, with allowance for the long-range magnetic interaction between the vortices through normal space, is described,[4] instead of Eq. (1), by the spatially nonlocal equation

$$\varphi_{tt}+\gamma\varphi_t+\sin\varphi=\frac{\partial}{\partial x}\int Q(x,z)\,\frac{\partial}{\partial z}\varphi(t,z)dz+j_0(x). \tag{28}$$

Here the source of current $j_0(x)$, localized primarily near the edges, is an external transport current [the integral of the first term on the right-hand side of Eq. (28) over the entire length of the junction is zero]. For a (formally) infinitely wide junction the kernel $Q(x,z)$ depends only on the difference of the arguments, and for in-line geometry has, in dimensionless variables, the form[4]

$$Q(x)=\frac{1}{\pi\alpha}K_0\left(\frac{|x|}{\alpha}\right)+\frac{\lambda}{\pi D}\int_0^\infty\frac{J_0(qx)dq}{\sigma^3[\sigma+\alpha q\,\coth(\sigma D/\lambda)]}, \tag{28a}$$

where $\sigma=\sigma(q)\equiv\sqrt{1+\alpha^2q^2}$. Here the parameter α is the ratio of the London penetration depth in the arms to the Josephson length λ_j.

Next, we restrict the discussion to the case in which $\alpha\ll 1$ [as in the model (1)], and the ratio of the thickness to the penetration depth is large, $D\gg\lambda$, but finite. We can then replace the first term in (28a) by a delta function. The integral in (28a), which we denote by $R(x)$, for $|x|\gg\alpha$ always has a power ''tail'':

$$R(x)\simeq\frac{1}{|x|}.$$

For $\varphi(t,x)$ far from the edges of a wide junction, Eq. (28) yields

$$\varphi_{tt}+\gamma\varphi_t+\sin\varphi=\varphi_{xx}+\frac{\lambda}{\pi D}j_{\rm nl}(t,x), \tag{29}$$

where

$$j_{\rm nl}(t,x)\equiv\frac{\partial}{\partial x}\int R(x-z)\,\frac{\partial}{\partial z}\varphi(t,z)dz$$

is the nonlocal contribution. The inequality $D\gg\lambda$ makes it possible to interpret this contribution as a weak perturbation to the local dynamics described by Eq. (1). The approximation specified by Eqs. (12) and (13) is still meaningful, but the balance equation (16) acquires on the right-hand side an additional term $(\lambda/\pi D)j_{\rm nl}\varphi_t$ (the equation remains exact, but the expression in parentheses on the left-hand side is

only the energy localized in the junction). Ignoring the fluctuations in the nonlocal contribution to the current density, i.e., performing the substitution

$$\langle j_{\rm nl}\varphi_t\rangle \simeq \langle j_{\rm nl}\rangle\Omega = \Omega\frac{\partial}{\partial x}\int R(x-z)h(z)dz,$$

we obtain at the following equation [instead of Eq. (22)]:

$$\frac{dk}{dx} = -\frac{k^2}{\pi}E(k)\left[\gamma\Omega - \frac{\lambda}{\pi D}\frac{\langle j_{\rm nl}\rangle}{f(k)}\right], \tag{30}$$

where the function $f(k)$ was introduced in (21).

Now we allow for the fact $f(k)>1$ for all $k>0$. A weak nonlocality may affect appreciably the current–voltage characteristic, when the second term in the square brackets in (30) is equal to the first. Equation (30) shows that this is possible only in the vicinity of the ''beak'' of the current–voltage characteristic, since only there is the factor $\langle j_{\rm nl}\rangle$ approximately proportional to the current I, while the voltage, according to (24), exponentially decreases with increasing current. In this region, $h(x)$ exhibits a logarithmic dependence on the coordinate. To estimate the corresponding characteristic voltage, we replace the factor $\langle j_{\rm nl}\rangle$ in (30) by its value averaged over the junction's width,

$$\overline{j_{\rm nl}} \equiv \frac{1}{L}\int\langle j_{\rm nl}\rangle dx \cong \frac{I}{L}\ln\frac{L}{2x_0}, \tag{30a}$$

and nullify the right-hand side of Eq. (30). We thus obtain an estimate for the characteristic value of the normal current, $I_n^{\rm cr}$, at which there is a crossover to the current–voltage characteristics behavior determined by nonlocal effects:

$$I_n^{\rm cr} \simeq \frac{4\lambda}{D}\frac{\ln(L/2x_0)}{\ln(4\pi D/\lambda)} \sim \frac{\lambda}{D}$$

(the corresponding characteristic value of voltage in dimensional form is of the order of $U_n^{\rm cr}\sim\lambda RI_0/D=\lambda\lambda_j RLj_c$).

For lower normal current values the logarithmic ''beak'' is replaced by a smoother I-vs-I_n dependence. We can assume on the basis of (30) and (30a) that the dependence is approximately linear with a slope of the order of D/λ, greater than the slope in the ohmic region (equal to unity). Since the penetration depth increases with temperature, this initial slope of the current–voltage characteristic must decrease with increasing temperature.

Let us examine the physical meaning of the asymptotic behavior (24) and the transition from it to the quasilinear behavior of current–voltage characteristics. As the voltage is lowered, the distance between the vortices increases and becomes comparable to the size of a free fluxon. Here the vortex–vortex interaction exponentially decreases with increasing distance but must balance the friction force. Since, as we have seen earlier, the distance is of the order of $1/n=2\pi/h\sim 4\pi/I$, and since the frictional force is proportional to Ω, it becomes an exponential function of the reverse current, which is reflected in (24). Here the inner part of the junction acts as a ''bottleneck,'' since the vortex concentration inside is lower. In the presence of a nonlocality, however, the vortices at junction edges induce in the junction's center an almost evenly distributed current $(\lambda/\pi D)j_{\rm nl}(t,x)$. This current generates in (29) a Lorentz force, which, by opposing friction, pushes the ''inner'' vortices to the center (local vortex–vortex interaction also does this), which is reflected in (30). For $I_n<I_n^{\rm cr}$ the nonlocality begins to play the crucial role, and so the nonlocality in this region cannot be considered a weak perturbation.

4.3. Boundary conditions and the supercurrent distribution

Let us see how the spatial distributions of current and field change as the boundary conditions (1a) are replaced by (3). We write the approximate solution to Eq. (1) in the form

$$\varphi(t,x) = \varphi_0(t,x) + a(t,x),$$

where $\varphi_0(t,x)$ stands for the solution of Eq. (1) under conditions (3), and the second term is a ''plasma excitation,'' which is unrelated to a change in topological charge, i.e., the number of vortices in the junction. We consider the fairly distant resistance region, where quantitatively the difference between (1a) and (3) is small. From (1a), (3), and (13) we then obtain the following boundary conditions for the plasma oscillations:

$$a_x\left(t,\pm\frac{L}{2}\right) = \pm\frac{I}{2} - \varphi_{0x}\left(t,\pm\frac{L}{2}\right) \approx -\frac{2}{I}\cos\left[\Omega t + \theta\left(\pm\frac{L}{2}\right)\right]. \tag{31}$$

Since the factor $2/I$ is small, we assume that $a(t,x)$ is a weak perturbation. From Eq. (1) it follows that the corresponding addition to the current density can be written as

$$\delta j(x) = \frac{d^2\langle a\rangle}{dx^2} = \langle\sin\varphi\rangle - \langle\sin\varphi_0\rangle \simeq \langle a\cos\varphi_0\rangle, \tag{32}$$

where we have allowed for the fact that because of its zero-charge nature the plasma addition must satisfy the following relation $\langle\partial a(t,x)/\partial t\rangle = 0$. On the other hand, the weakness of the perturbation makes it possible to linearize Eq. (1) in this perturbation:

$$a_{tt} + \gamma a_t = a_{xx} - a\cos\varphi_0. \tag{33}$$

From the standpoint of this equation, the resistance modes with a double period $2(2\pi/\Omega)$ observed in Ref. 2 can be explained by parametric generation of subharmonics of Josephson frequency caused by the parametric term $a\cos\varphi_0 \simeq a\cos[\Omega t+\theta(x)]$. However, under the condition (2), such modes and the stochastic mode (if they exist at all) differ very little from the simplest mode with a period $2\pi/\Omega$ and are practically unidentifiable. This is understandable, since the parametric effect is caused by the ''bulk'' interaction of fluxons and the plasma wave. This interaction is weak because of the large difference between the velocity of the plasma wave (close to unity) and the fluxon velocity [which, if the condition (2) is satisfied, is less than unity almost everywhere).

This fact makes it possible to discard the parametric term in (33) and to replace $a(t,x)$ on the right side of Eq. (32) by the solution of the elementary wave equation $a_{tt}+\gamma a_t = a_{xx}$ with the boundary conditions (31). Moreover,

according to the above analysis, in the ohmic region considered here we can employ the following approximation:

$$\varphi_0 \cong \Omega t + \theta(x), \quad \theta(x) \approx \frac{I_s|x| + \gamma\Omega x^2}{2} + \text{const.} \tag{34}$$

As a result, after doing standard calculations and averaging over time, we obtain from (32) the following redistribution of the supercurrent:

$$\delta j(x) \approx \mathrm{Re}\left\{\left[IS \sinh\frac{SL}{2}\right]^{-1} \cosh Sx \times \exp\left[i\left(\theta\left(\frac{L}{2}\right) - \theta(x)\right)\right]\right\}, \tag{35}$$

where we have introduced the complex–valued wave number $S=(i\gamma\Omega-\Omega^2)^{1/2}$.

For (34) and (35) we see that the ''plasma waves'' lead to spatial oscillations of the time–averaged current density. Here both the amplitude and the spatial frequency of these oscillations building up from the middle to the edges of the junction. This pattern was observed in the numerical experiments mentioned above.

5. CONCLUSIONS

We have analyzed an approximate analytic theory of the stationary resistance state of Josephson junctions with a low normal resistance. Although our results pertain primarily to the specific mathematical model of Eq. (1), they can easily be generalized to the case of nonlinear friction. In addition, our approach may prove useful in analyzing the interaction of a resistance state with high–frequency electromagnetic radiation and constant external fields. In particular, interesting phenomena occur and find their reflection in the incremental resistance of the junction (the derivative of the current–voltage characteristic) in the presence of a spatially modulated, static magnetic field.[5–7]

One possible area of application of our results, in addition to artificial junctions, is presented by naturally emerging, extended weak bonds in granular and polycrystalline high-T_c superconductors. For instance, the films studied in Refs. 5–7 had a resistance $R \simeq 1\ \Omega$ per weak bond. At the same time, estimates of λ and D for these films yield $D/\lambda \approx 3-10$, so that the critical resistance R_0 defined earlier amounts to 3–30 Ω. Thus $R < R_0$, and the condition (2) of viscous magnetic−flux transport is met [actually, the strong inequality in (2) is unnecessary since, as we have seen, in the nonballistic mode the real vortex velocity was found to be lower than the maximum velocity, so that the mean free path of fluxons is shorter than the one adopted in deriving (2)].

I wish to thank A. K. Asadov, Yu. V. Medvedev, and P. N. Mikheenko for the useful discussions.

[1] I. O. Kulik and I. K. Yanson, *Josephson Effect in Superconducting Tunnel Structures* [in Russian], Nauka, Moscow (1970).
[2] A. E. Filippov, Yu. E. Kuzovlev, and T. K. Soboleva, Phys. Lett. A **183**, 123 (1993).
[3] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, National Bureau of Standards Applied Mathematics Series 55, Washington, D.C. (1964).
[4] Yu. E. Kuzovlev and A. I. Lomtev, Zh. Éksp. Teor. Fiz. **111**, 1803 (1997) [JETP **84**, 986 (1997)].
[5] Yu. M. Ivanchenko, Yu. E. Kuzovlev, P. N. Mikheenko, and Ya. I. Yuzhelevskiĭ, Sverkhprovodimost' **4**, 950 (1991) [Superconductivity **4**, 859 (1991)].
[6] Ya. I. Yuzhelevskiĭ, Yu. E. Kuzovlev, and P. N. Mikheenko, Fiz. Nizk. Temp. **17**, 1377 (1991) [Sov. J. Low Temp. Phys. **17**, 741 (1991)].
[7] P. N. Mikheenko and Ya. I. Yuzhelevskiĭ, Physica C **217**, 97 (1993).

Translated by Eugene Yankovsky
Edited by S. J. Amoretty

Suppression of the fractal conductivity channel and superlocalization effects in porous *a*-Si:H

A. I. Yakimov, N. P. Stepina, A. V. Dvurechenskiĭ, and L. A. Shcherbakova

Institute of Semiconductor Physics, Russian Academy of Sciences, Siberian Branch, 630090 Novosibirsk, Russia

(Submitted 15 December 1996)
Zh. Éksp. Teor. Fiz. **112**, 926–935 (September 1997)

The temperature dependence of the hopping conductivity and the relaxation kinetics of the transient current in porous amorphous silicon are investigated after treatment in a hydrogen plasma at 200 °C. It is discovered that posthydrogenation of the material increases the dimension of the conducting channel from 2.5 to 3, while suppressing and slowing the relaxation of the transient current. The results obtained are attributed to passivation of the electrically active dangling bonds on the pore surface by hydrogen. It is concluded that electron transport in porous amorphous silicon in the temperature range $T>T^*$, where T^* lies in the range 130–270 K and depends on the density of states, takes place between superlocalized states of the internal surface, which is enriched with dangling bonds and acts as a fractal percolation system. When the temperature is lowered below T^*, a transition to one-dimensional hopping conduction in the bulk silicon regions occurs. © *1997 American Institute of Physics.* [S1063-7761(97)01209-2]

1. INTRODUCTION

Porous silicon is formed as a result of the electrolysis of monolithic (compact) silicon in a solution of hydrofluoric acid. The electrolyte flows among the sites of lowest electrical resistance and etches out a system of pores in the material in the form of an infinite cluster. The infinite cluster is fractal at the percolation threshold and has the dimension $D\approx 2.5$ in three-dimensional Euclidean space. This result is a consequence of percolation theory.[1] The fractal nature of the remaining silicon skeleton is not a trivial point. Its fractal dimension should, at least, not coincide with the dimension of the pore system or, therefore, with the dimension of the infinite cluster in percolation theory due to the presence of bulk silicon regions. Nevertheless, investigations of conduction in undoped porous amorphous silicon prepared either by ion implantation[2] or by sputtering in an ultrahigh vacuum[3] showed that in some cases charge transport is effected by electron hopping in a fractal medium of dimension close to 2.5. Figure 1 presents the dependence of the effective dimension of the conducting channel on the porosity in porous amorphous silicon obtained on the basis of different layers of amorphous silicon (*a*-Si). Data from Refs. 2 and 3 were used here. The arrow points to the dimension of the percolation cluster in percolation theory.

The temperature dependence of the hopping conductivity $\sigma(T)$ in amorphous materials is described in the general case by the expression

$$\sigma(T)=\sigma_0 \exp\left[-(T_0/T)^x\right], \tag{1}$$

where σ_0 and T_0 are constants having the dimensions of electrical conductivity and temperature, respectively, and the exponent x is determined only by the dimension D of the system [if the density of states $g(E)$ does not depend on the energy]. The literature describes several mechanisms of hopping conduction, which lead to different values of x:

$$x=1/(1+D) \tag{2}$$

for the familiar Mott law in the absence of Coulomb correlations[4] (in particular, $x=0.25$ for three-dimensional systems); and

$$x=1 \tag{3}$$

for one-dimensional hopping conduction;[5–7]

$$x=\zeta/(\zeta+D) \tag{4}$$

for charge transfer in a medium of fractal dimension D.[8] Here ζ characterizes the features of the decay of the wave function of an electron localized on a fractal. In fact, the anomalous nature of diffusion in a fractal medium leads to the superlocalization of such an electron.[9,10] At large distances its wave function begins to decay more rapidly than Anderson decay:

$$\Psi(r)\propto \exp\left[-(r/a_s)^{\zeta}\right]. \tag{5}$$

For an infinite cluster at the percolation threshold the theory gives $\zeta=1.9$ (Ref. 9) and $x=0.43$ (the three-sevenths law[8]). Similar values of x were observed in the experiments in Refs. 2 and 3. The question of the origin of the fractal conduction channel in porous *a*-Si was still open when the present work was undertaken.

2. FORMULATION OF THE PROBLEM

Porous silicon is known to be characterized by numerous phase boundaries and an enormous internal surface. Its specific area can amount to 200–600 m^2/cm^3.[11] The internal surface of the silicon skeleton is a ''cast'' of the percolation cluster forming the system of pores and should, therefore, have a fractal dimension close to 2.5.

In undoped amorphous silicon, dangling bonds, which produce a peak of deep localized states with a density as high as 10^{20} $eV^{-1}\,cm^{-3}$ in the mobility gap, are responsible for hopping charge transfer. It is natural to expect that the internal surface of porous amorphous silicon will be richer in

 1063-7761/97/090501-06$10.00

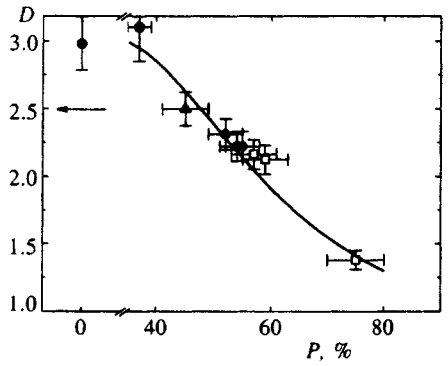

FIG. 1. Dependence of the effective dimension of the conducting channel on the porosity in layers of porous amorphous silicon, according to the data in Ref. 2 (●), Ref. 1 (□), and the present work (▲). The dimension was determined using Eq. (2) for $P<40\%$ and Eq. (4) for $P>40\%$; P is the porosity.

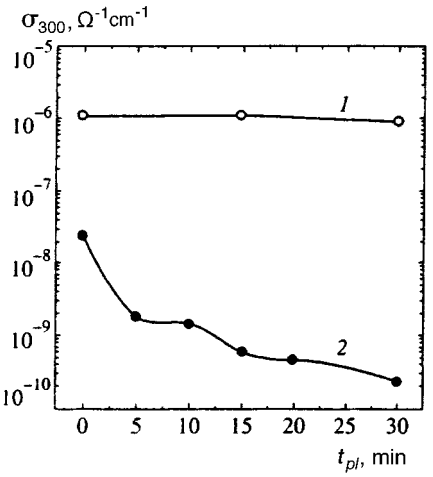

FIG. 2. Dependence of the conductivity at room temperature on the annealing time in a hydrogen plasma for samples of monolithic a-Si (*1*) and porous a-Si (*2*). The annealing temperature was 200 °C.

dangling bonds than the bulk silicon regions; therefore, its contribution to the conductivity can be far more significant than the contribution of the bulk. In addition, the superlocalized electron states should also be located specifically on the fractal surface. Thus, there are two parallel conduction channels: a surface channel between superlocalized states of the fractal structure and a bulk channel of normal hopping transport in the a-Si remaining after the etching. In structures with a small porosity P the bulk conduction channel dominates, and the effective dimension of the system is close to 3 (Fig. 1). As P increases the relative contribution of the bulk decreases, and the dimension of the conducting channel is determined by the fractal structure of the surface (the porosity is then about 45%). In structures with a porosity of order 70–80% the silicon regions degenerate into thin one-dimensional channels. Smoothing of the pore surface apparently then occurs, and its contribution to the conductivity of the system decreases. This result is in good agreement with the decrease in specific surface as the porosity of porous crystalline silicon increases, discovered in Ref. 12.

The existence of different channels can be revealed experimentally by suppressing one of them. This goal is achieved, for example, by posthydrogenation of the structures in a hydrogen plasma. Atomic hydrogen is known[13] to saturate dangling bonds by directly reacting with them and forming a Si–H chemical bond, whose energy level lies outside the mobility gap. By selecting the treatment conditions so that H^+ will not penetrate far into the bulk of a-Si and will interact only with states near the pore surface, we can suppress the surface conductivity. In addition, the increase in the thickness of the near-surface carrier-depleted layer as a consequence of the decrease in the density of states stimulated by the interaction with hydrogen should lead to the blocking of transport near the surface.

3. EXPERIMENTAL METHOD

The experiment was carried out on structures with layers of porous amorphous silicon, whose porosity determined by the method described in Ref. 3 was about 45%. Specifically in these samples the dimension of the conducting channel was close to the dimension of the fractal percolation cluster (Fig. 1). The layers of amorphous silicon were formed on degenerate substrates of crystalline n-type silicon with a resistivity of 0.002 Ω cm by implanting Ge^+ ions in a dose $\Phi=2\times10^{15}$ cm^{-2}. The energy of the ions was $E=400$ keV. The electrolysis was carried out in a 42% HF:H_2O:C_3H_7OH (1:1:2) solution. The anode current was $j=15$ mA/cm^2, and the etching time was $t=20$ s. Detailed information on the microscopic structure of the porous amorphous silicon, the features of the preparation of the contacts, and the conductivity measurements can be found in Refs. 2, 3, and 14.

Passivation of the dangling bonds was achieved by treating the structures in a hydrogen plasma generated in an hf glow discharge (the frequency was 13.5 MHz). The samples were placed outside the discharge zone, and hydrogen ions were supplied to the sample as a result of the motion of H^+ under the action of the pressure drop from the discharge zone to the sample surface. This made it possible to avoid the problems associated with contamination of the sample surface by the chamber. The hydrogenation temperature was $T=200$ °C. The diffusion coefficient of hydrogen from the plasma into the ion-implanted layers of a-Si at such a temperature is $\approx10^{-16}$ cm^2/s,[15] and the penetration depth after a time on the order of 10 min is $l_D\approx20$ Å. Such treatment should not have a significant influence on defects in the silicon bulk, and the reaction of hydrogen with the dangling bonds should take place mainly on the pore surface. In addition, the high concentration of surface defects that act as traps for hydrogen should prevent its penetration into the silicon bulk. These circumstances permit precise variation of the ratio between the surface and bulk charge-transfer channels. Figure 2 presents plots of the dependence of the room-temperature conductivity σ_{300} on the plasma-treatment time $\tau_{\rm pl}$ in monolithic a-Si and porous amorphous silicon. It is seen that the plasma-annealing procedure scarcely alters σ_{300} for the monolithic material, while in porous silicon its value decreases by more than two orders of magnitude in the range $\tau_{\rm pl}=0$–30 min.

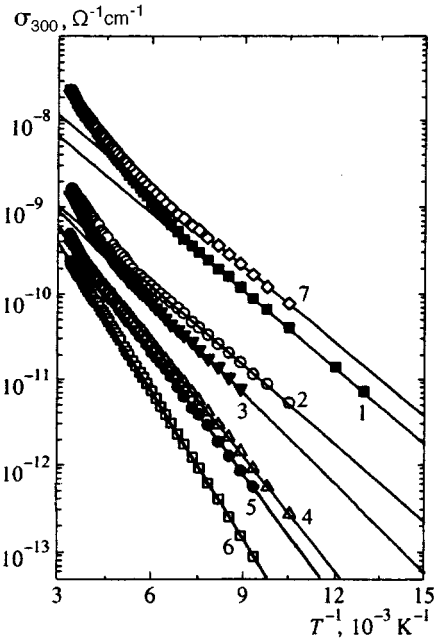

FIG. 3. Temperature dependence of the conductivity in samples of porous amorphous silicon after various hydrogenation times τ_{pl}: *1*—original sample (τ_{pl}=0 min), *2*—5 min, *3*—10 min, *4*—15 min, *5*—20 min, *6*—30 min. Curve *7* corresponds to porous amorphous silicon that was treated in the plasma for 20 min and then annealed in a vacuum at T=400 °C for 25 min.

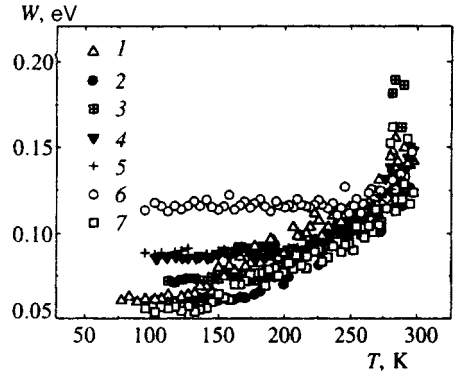

FIG. 4. Temperature dependence of the activation energy for conduction in porous amorphous silicon obtained by direct differentiation of experimental $\sigma(t)$ curves. Plasma treatment time: *1*—0 min, *2*—5 min, *3*—10 min, *4*—15 min, *5*—20 min, *6*—30 min. Curve *7* corresponds to porous amorphous silicon treated in the plasma for 20 min and then annealed in a vacuum at T=400 °C for 25 min.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Temperature dependence of the conductivity

The temperature dependence of *a*-Si layers not subjected to electrolysis follows Mott's law (2) with $x=0.25$ in the temperature range 77–300 K and varies weakly as the hydrogenation time is increased. The density of localized states extracted from the Mott dependence of $\sigma(T)$ decreases from 2.2×10^{20} eV$^{-1}\cdot$cm^{-3} for the undoped material to 1.9×10^{20} eV$^{-1}\cdot$cm^{-3} for τ_{pl}=30 min.

Figure 3 presents the temperature dependence of the conductivity of layers of porous amorphous silicon for various values of τ_{pl} in $\log\sigma$ versus T^{-1} coordinates. At low temperatures a linear dependence, which corresponds to law (3) and is associated with one-dimensional hopping transport,[2,16] is observed in all cases. Since the pores have a predominant orientation perpendicular to the sample surface, the increase in the hopping distance as the temperature is lowered should lead to one-dimensionalization of the electron trajectories along the direction demarcated by the pore walls.[2] Approximation of the low-temperature part of the $\sigma(T)$ curve by the relation

$$\sigma(T) \propto \exp\left(-\frac{1}{2g_1 akT}\right), \tag{6}$$

which was obtained theoretically for the one-dimensional case,[6,7] makes it possible to determine the one-dimensional density of states g_1. Here $a\approx3$ Å is the localization radius, and k is the Boltzmann constant. The results are presented in Table I.

The course of the $\sigma(T)$ curve is conveniently investigated by analyzing the local activation energy of the conduction process,[17,18] $W(T)=-\partial(\ln\sigma)/\partial(1/kT)$. It is easily seen from Eq. (1) that

$$W(T)=x(kT_0)^x(kT)^{1-x}. \tag{7}$$

Figure 4 shows plots of the temperature dependence of the activation energy, which were obtained by direct differentiation of the experimental set of $\sigma(T)$ points. Two conduction regimes are clearly observed for all the samples: In the low-temperature range $T<T^*$ the value of W does not depend on T; at $T>T^*$ the activation energy increases monotonically

TABLE I. Parameters of the temperature dependence of the conductivity.

τ_{pl}, min	σ_0, Ω^{-1}cm^{-1}	T_0, K	x	D[a]	g_1, eV^{-1}cm^{-1}
0	$(6.8\pm0.7)\times10^{-5}$	$(6.5\pm0.9)\times10^{4}$	0.43±0.01	2.51±0.11	$(2.8\pm0.3)\pm10^{8}$
5	$(4.7\pm0.7)\times10^{-6}$	$(1.0\perp0.2)\times10^{5}$	0.40±0.01	2.83±0.12	$(2.5\pm0.2)\times10^{8}$
10	$(7.9\pm1.7)\times10^{-6}$	$(1.3\pm0.4)\times10^{5}$	0.39⊥0.02	2.93±0.12	$(2.3\pm0.1)\times10^{8}$
15	$(1.3\pm0.7)\times10^{-2}$	$(3.7\perp1.2)\times10^{7}$	0.25±0.03	3.00±0.3	$(2.0\pm0.1)\times10^{8}$
20	$(6.2\pm0.7)\times10^{-2}$	$(4.5\pm0.7)\times10^{7}$	0.25±0.03	3.00±0.3	$(1.8\pm0.1)\times10^{8}$
30	0.12±0.07	$(4.8\pm0.9)\times10^{7}$	0.26±0.10	3.00±1.1	$(1.4\pm0.1)\times10^{8}$
20[b]	$(4.5\pm0.5)\times10^{-3}$	$(1.9\pm0.3)\times10^{5}$	0.39±0.02	2.93±0.12	$(3.0\pm0.2)\times10^{8}$

[a]The effective dimension D of the conducting channel was determined using Eq. (4) for a plasma treatment time of 0–10 min and Eq. (2) for a time of 15–30 min.

[b]After plasma treatment the sample was subjected to heating in a vacuum with a residual pressure of 10^{-4} Pa at 400 °C for 25 min.

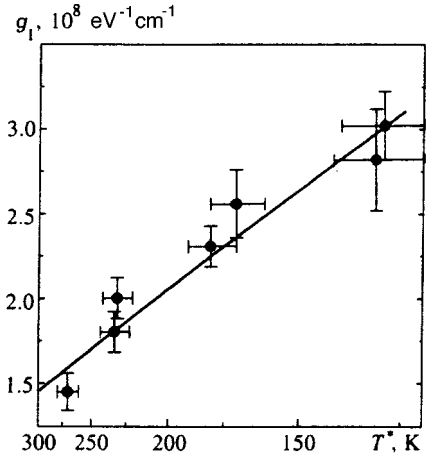

FIG. 5. One-dimensional density of states determined at $T<T^*$ on the basis of an analysis of the Arrhenius law (6) vs the transition temperature T^*.

with increasing temperature. The temperature below which an Arrhenius law is observed, which we interpret as the temperature of the transition to a quasi-one-dimensional hopping conduction regime,[2,3] depends on the plasma treatment time. One-dimensionalization of the electron trajectories in porous silicon takes place at the temperature corresponding to the condition of equality between the hopping distance R and the diameter d of the silicon filaments. In a regime of hopping conduction among states near the Fermi level, the hopping distance is determined by the temperature and the density of localized states: $R(T)\propto 1/(Tg)^x$. This means that the parameters g_1 and T^* should be related by the expression $d\approx A/(T^*g_1)^x$, where A is a parameter which reconciles the units of measure. Hence, $g_1\approx(A/d)^{1/x}/T^*$. Thus, if it is correct to attribute the change in the character of $\sigma(T)$ with decreasing temperature to a transition to one-dimensional hopping transport, the density of states should be a linear function of the reciprocal temperature of the transition. Figure 5 shows the dependence of the one-dimensional density of states on T^*. The plot was constructed by analyzing the data for samples treated in the plasma for different times. Fitting the experimental dependence of $g_1(T^*)$ to the power function $g_1(T^*)=B/T^{*\beta}$ permitted determination of the parameters which provide the best fit: $B=(2\pm1)\cdot10^{10}$ K$^{0.9}\cdot$eV$^{-1}\cdot$cm^{-1} and $\beta=0.9\pm0.1$ (the solid line in Fig. 5). The specific dependence of the maximum hopping distance on the temperature and the density of states in one-dimensional systems was obtained by Raĭkh and Ruzin by optimizing the form of the ''cuts'' in a one-dimensional chain[6] and by Zvyagin on the basis of percolation theory.[7] The former model gives, $x=1$ and $A=1/(2k)$, and Zvyagin's model gives $x=1/2$ and $A=(2a/k)^{1/2}$. Using the relation between A and B in the form $B=(A/d)^{1/x}$, we can estimate the characteristic diameter of the silicon filaments. It was found that both theoretical approaches give approximately the same value $d\approx45\pm20$ Å, which is consistent with the data from high-resolution electron microscopy.[14]

Table I presents the results of approximating the experimental data for $T>T^*$ by Eq. (1) using the least-squares method. The fitting parameters were σ_0, T_0, and x. The original sample of porous silicon had $x=0.43$, which accurately corresponds to a fractal dimension $D=2.51$ of the conducting channel. As a result of hydrogenation, the value of x decreases, and the value of the effective dimension obtained from Eq. (4) increases. The deviation of D from 2.51 in the structures treated in the plasma is a result of the contribution of bulk three-dimensional conduction in the silicon remaining after electrolysis. The samples treated for 15 and 20 min had $x\approx0.25$. This means that the conductivity follows Mott's law (2) in a Euclidean space with a dimension equal to 3. Apparently, in these cases passivation of the surface states by hydrogen completely suppresses the fractal conduction channel. A similar picture is observed for $\tau_{\rm pl}=30$ min, although the accuracy of the determination of x in this case is low because of the narrow width of the temperature range 300 K$>T>T^*$.

4.2. Anomalous relaxation

A direct consequence of the existence of superlocalized electron states like (5) on the fractal surface in porous amorphous silicon is the presence of long-lived excitations of the electronic system. It was discovered in Refs. 3 and 14 that application of a voltage step or pulse to a porous layer yields a time-dependent electric current $I(t)$, which slowly decays to its stationary value I_s. A detailed analysis of the current relaxation kinetics showed that they are anomalous (non-Debye), i.e., they are not described by an exponential dependence with a single decay time, and that the characteristic relaxation times (of the order of 10^3 s) significantly exceed the Maxwell time in this material. The appearance of the transient current is caused by the injection of carriers into localized states above the Fermi level followed by thermalization to a quasiequilibrium state.[14] The downward movement of the electrons along the energy scale is accomplished by tunneling between localization centers with phonon emission. The occurrence of superlocalization strongly slows the relaxation kinetics and leads to the existence of transient currents for hundreds of seconds. Elimination of the superlocalized states on the surface and the associated fractal conduction by passivation in a hydrogen plasma should lead to the disappearance of the long-term anomalous relaxation in porous amorphous silicon.

Figure 6 shows plots of the time dependence of the transient component $\Delta I(t)=I(t)-I_s$ of the current in layers of porous silicon after different treatment times $\tau_{pl}=0-15$ s in the plasma. In the sample held in the plasma for 30 min the time dependence of the current could not be detected to within the sensitivity of the measuring instrument (~1 pA). The system was excited by applying a voltage step with a height of 60 mV. The high resistance of the structures and, therefore, the long times needed for establishment of the transient processes in the circuit call for the performance of correct measurements only at times $t>1-5$ s. Therefore, the measurements of the relaxation kinetics were begun only 5 s after the bias voltage was switched on.

Figure 6 shows that the plasma treatment results, first, in slowing of the relaxation process and, second, in a decrease in the magnitude of the transient current. Both these findings

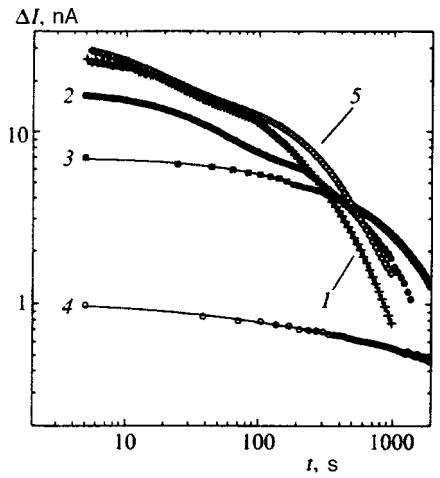

FIG. 6. Kinetics of the relaxation of the transient current at room temperature excited by applying a voltage step with a height of 60 mV to the sample. The rise time of the leading edge of the step was 10 ms. The numbers of the curves correspond to different hydrogenation times: *1*–0 min, *2*—5 min, *3*—10 min, *4*—15 min. Curve *5* corresponds to porous amorphous silicon treated in the plasma for 20 min and then annealed in a vacuum at $T=400$ °C for 25 min.

are easily attributed to suppression of the fractal conduction channel as a consequence of a decrease in the concentration of superlocalized states on the surface. Since the rates of the tunneling transitions exhibit a strong exponential dependence on the distance between the states and, consequently, on their concentration, elimination of the dangling bonds should impede the establishment of a stationary state in the system. The magnitude of the excess current in the system is determined by the number of carriers injected into states lying above the Fermi level that are free at equilibrium. The decrease in the concentration of these states in the hydrogenated material naturally leads to suppression of the transient injection currents.

Of course, the relaxation channel involving ordinary (not superlocalized) bulk silicon states then persists. However, the characteristic times of such a process in amorphous materials are fairly short[19] (10^{-6}–10^{-4} s). Therefore, it is not observed under the conditions of the present experiment.

4.3. Restoration of the fractal properties of porous amorphous silicon

It should be noted that plasma treatments of semiconductor materials are often employed to etch the surfaces of samples.[20] In this case the action of the hydrogen plasma reduces not to chemical passivation of the dangling bonds on the pore surface, but to its ''smoothing'' either directly by mechanically removing bumps or by activating chemical reactions of silicon with fluorine, which is present in porous amorphous silicon in large amounts.[3] Etching a surface can lead to the irreversible loss of its fractal properties and consequently to suppression of the fractal conduction channel and the long-lived transient currents.

To elucidate the mechanism of the plasma-stimulated disappearance of the fractal properties in porous amorphous silicon, we measured the conductivity and the transient current in a sample subjected to plasma treatment for 20 min followed by heating in a vacuum with a residual pressure of 10^{-4} Pa at 400 °C for 25 min. When *a*-Si:H films are heated to such temperatures, hydrogen leaves the material, and the density of the dangling bonds increases.[21,22] It is seen from Figs. 2 and 3 and Table I that annealing results in practically complete restoration of the conductivity, its temperature dependence, and the current relaxation kinetics. This finding is evidence that the dominant contribution to suppression of the fractal properties of porous silicon during plasma treatment is the reversible passivation of surface states by hydrogen.

5. CONCLUSIONS

The measurements performed here of the temperature dependence of the conductivity and the kinetics of the decay of the transient current excited by applying a voltage step in porous amorphous silicon that was subjected to hydrogenation in a hydrogen plasma have made it possible to establish that the special features of electron transport in this material (the deviation of the conductivity from the Mott behavior and the existence of long-lived transient currents) are caused by the presence of an internal fractal surface with a dimension ≈ 2.5, which is enriched with dangling silicon bonds. Passivation of these bonds by hydrogen increases the effective dimension of the system from 2.5 to 3 as a consequence of the mixing of the fractal and Euclidean conduction channels, while suppressing and slowing the anomalous relaxation kinetics of the surface currents. The experiments performed show that plasma treatment, like electrolytic etching, permits adjustment of the dimension of the conducting channel in porous silicon.

The conclusion that the competition between the surface and bulk processes plays a decisive role in shaping the charge transport process in silicon is confirmed by the results of the investigations on porous layers of *a*-Si doped with manganese.[23] The study of the conductivity in $a\text{-}Si_{1-c}Mn_c$ structures ($c=0.04$ and 0.07) subjected to electrolysis under conditions similar to the conditions of the present work revealed the absence of fractal properties in the temperature range 20–300 K. The variation of the conductivity with the temperature was described by Mott's law with $x=1/4$, which was replaced by the Arrhenius law corresponding to one-dimensional transport as T is lowered. In addition, long-lived transient currents were not observed in this material. Such behavior is caused by the high density of Mn impurity states in the bulk silicon regions ($\sim 10^{21}$ $eV^{-1}\cdot cm^{-3}$), which leads to dominance of the Euclidean conduction channel.

We express our sincerest thanks to É. M. Baskin for numerous useful discussion. This work was performed with partial support from the international association INTAS (Grant 94-4435), the ''Current Research Areas in the Physics of Condensed Media'' State Program (in the area of ''Surface Atomic Structures,'' Grant 97-3.21), and the ''Fundamental Research in the Field of Chemical Technologies'' State Problem for Supporting Universities in Russia (Grant 3N-254-97).

[1] B. I. Shklovskiĭ and A. L. Efros, *Electronic Properties of Doped Semiconductors*, Springer-Verlag, Berlin (1984) [Russ. original, Nauka, Moscow (1979, p. 126)].
[2] A. I. Yakimov, N. P. Stepina, A. V. Dvurechenskii, and L. A. Scherbakova, Physica B **205**, 298 (1995).
[3] A. I. Yakimov, N. P. Stepina, A. V. Dvurechenskiĭ, L. A. Shcherbakova, and A. I. Nikiforov, Zh. Éksp. Teor. Fiz. **110**, 322 (1996) [JETP **83**, 174 (1996)].
[4] N. F. Mott, Philos. Mag. **19**, 835 (1969).
[5] J. Kurkijarvi, Phys. Rev. B **8**, 922 (1973).
[6] M. E. Raĭkh and I. M. Ruzin, Zh. Éksp. Teor. Fiz. **95**, 1113 (1989) [Sov. Phys. JETP **68**, 642 (1989)].
[7] I. P. Zvyagin, Zh. Éksp. Teor. Fiz. **107**, 175 (1995) [JETP **80**, 93 (1995)].
[8] G. Deutscher, Y.-E. Lévy, and B. Soullard, Europhys. Lett. **4**, 577 (1987).
[9] Y.-E. Lévy and B. Soullard, Europhys. Lett. **4**, 233 (1987).
[10] A. Aharony and A. Brooks Harris, Physica A **163**, 38 (1990).
[11] R. Herino, G. Bomchil, K. Barla *et al.*, J. Electrochem. Soc. **134**, 1994 (1987).
[12] A. Halimaoui, Surf. Sci. Lett. **306**, L550 (1994).
[13] D. Kaplan, in *The Physics of Hydrogenated Amorphous Silicon*, J. D. Joannopoulos and G. Lucovsky (Eds.), Springer-Verlag, Berlin–New York (1984) [Russ. transl., Mir, Moscow (1987), p. 235].
[14] A. I. Yakimov, N. P. Stepina, A. V. Dvurechenskii, L. A. Scherbakova, and A. I. Nikiforov, Phys. Rev. B (in press).
[15] A. V. Dvurechenskiĭ, I. A. Ryazantsev, and L. S. Smirnov, Fiz. Tekh. Poluprovodn. **16**, 621 (1982) [Sov. Phys. Semicond. **16**, 400 (1982)].
[16] A. I. Yakimov, N. P. Stepina, A. V. Dvurechenskii *et al.*, Philos. Mag. Lett. **73**, 17 (1996).
[17] A. G. Zavrodskiĭ and I. S. Shlimak, Fiz. Tekh. Poluprovodn. **9**, 587 (1975) [Sov. Phys. Semicond. **9**, 391 (1975)].
[18] R. M. Hill, in *Proceedings of the 7th International Conference on Amorphous and Liquid Semiconductors*, W. E. Spear (Ed.), University of Edinburgh, Edinburgh (1977), p. 229.
[19] H. Scher and E. W. Montroll, Phys. Rev. B **12**, 2455 (1975).
[20] Ku-Hyun Hwang, Euijoon Yoon, Ki-Woong Whang, and Jeong Yong Lee, Appl. Phys. Lett. **67**, 3590 (1995).
[21] M. N. Brodsky, M. A. Frish, J. F. Ziegler, and W. A. Lanford, Appl. Phys. Lett. **30**, 561 (1977).
[22] J. I. Pankove and D. E. Carlson, Appl. Phys. Lett. **31**, 450 (1977).
[23] A. I. Yakimov, A. V. Dvurechenskii, N. P. Stepina *et al.*, J. Phys.: Condens. Matter **9**, 889 (1997).

Translated by P. Shelnitz

Vortex lattice kinetics and the electrodynamics of rigid superconductors

S. E. Savel'ev and L. M. Fisher

All-Russian Electrotechnical Institute, 111250 Moscow, Russia

V. A. Yampol'skiĭ

Institute of Radio Physics and Electronics, Ukrainian National Academy of Sciences, 310085 Khar'kov, Ukraine

(Submitted 27 December 1996)
Zh. Éksp. Teor. Fiz. **112**, 936–952 (September 1997)

A kinetic description is proposed for the dynamics of vortex lattices in rigid superconductors located in a magnetic field whose direction varies. The collision integral in the kinetic equation for the vortex density includes the crossing and successive regeneration of vortex filaments. The second equation of the theory expresses force balance: the equality of the magnetic force to the pinning force. It is shown that the magnetic force contains a collective term which depends on the vortex distribution function. The model is used as a basis for the electrodynamic equations of the critical state for the case of crossed magnetic fields. The transition from the proposed theory to the previously developed two-velocity quasihydrodynamic model is discussed. © *1997 American Institute of Physics.* [S1063-7761(97)01309-7]

1. INTRODUCTION

The static and quasistatic electromagnetic properties of rigid superconductors are usually discussed in terms of a critical state model.[1] This model makes it possible to calculate the static magnetization of a superconductor, hysteresis loops, and magnetic reversal losses for cases where the external magnetic field varies in magnitude, but its direction remains fixed. The model equation for the distribution of the magnetic induction has been justified in terms of a balance of forces acting on Abrikosov vortices. It is assumed that the magnetic force acting on a vortex is compensated by pinning forces related to the interactions of the vortices with various inhomogeneities of the crystal lattice. Because of pinning, the critical current density J_c turns out to be nonzero, and the penetration of the magnetic field is shielded by the superconducting currents.

The distribution of the magnetic induction is much more complicated if the external magnetic field contains several components and its orientation varies with time. A number of interesting electromagnetic phenomena have been observed under just these conditions.[2–5] The basic conclusion that follows from these experiments is that in all those spatial regions into which a variable electromagnetic field (even a weak one) penetrates, a constant current is displaced. In particular, it has been shown[4] that the static magnetic moment of a sample vanishes completely if a transverse variable magnetic field penetrates the superconductor in its entirety.

Calculating the electromagnetic properties of superconductors in an external magnetic field with a variable direction requires additional information about the direction of the critical current density vector $\mathbf{J}_c$. Until recently, problems of this sort were examined in terms of two different approaches. The first was developed in the papers of Clem and Gonzalez.[6–8] The most important feature of this approach was that it included the crossing of vortex filaments (flux-line cutting), which plays a fundamental role under these conditions. This phenomenon and the effects associated with it have been studied in detail.[9–12] Unfortunately, the influence of the crossing of vortices on the condition for force balance in a vortex system are not taken into account in the Clem–Gonzalez equations. In addition, these equations are in striking disagreement with these experiments.[2–5] The other approach, proposed in Refs. 2–4 for interpreting the experiments, uses a generalized equation for the critical state,

$$\nabla \cdot \mathbf{B} = \frac{4\pi}{c} J_c \frac{\mathbf{E}}{E}, \tag{1}$$

where $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic induction, and c is the speed of light. In other words, it is assumed that, as in the case of Ohm's law, the current density is directed along the electric field. This equation has been used to explain a number of nonlinear interactions of electromagnetic waves.[13,14] Unfortunately, like the Clem–Gonzalez model, Eq. (1) does not have an adequate physical justification.

The phenomena of pinning and the crossing of vortex filaments have been rigorously taken into account by Voloshin *et al.*,[15] who describe the dynamics of the vortical system in the framework of a two-velocity hydrodynamic model. One of the velocities describes the motion of the vortical lattice as a whole, while the other is associated with the relative motion of two intersecting sublattices. The electrodynamic equations can be derived by minimizing the Gibbs free energy. The derived equations include the model electrodynamic equations proposed in Refs. 6–8 and 2–4 as special cases.

The entire set of experimental data from these papers finds a natural explanation in the two-velocity hydrodynamic model. This model, however, can only pretend to a qualitative description, since the actual distribution of the vortices with respect to their angle of inclination ϑ relative to a given axis is replaced by just two vortical sublattices. In fact, the vortices within each physically small volume of a supercon-

ductor have different angles of inclination, so that a rigorous description of the vortical system must be in terms of an angular distribution function and employ a kinetic equation. The collision integral in the kinetic equation must provide a faithful description of the phenomenon of vortex crossing. Furthermore, the Gibbs free energy minimization proposed in Ref. 15, which serves as a basis for deriving the electrodynamic equations, requires more rigorous justification, including an analysis of the force balance condition. The influence of vortex crossing on force balance must be correctly accounted for in this analysis.

The purpose of the present paper is to construct a kinetic theory for the vortical system in a superconductor located in an external magnetic field $\mathbf{H}_0(t)$ whose direction is variable. Naturally, such a theory can be constructed only by using a number of reasonable assumptions and restrictions. Above all, here we assume that $H_0 \geqslant H_{c1}$. This makes it possible to neglect the distinction between the magnetic induction and the thermodynamic magnetic field (here H_{c1} is the lower critical magnetic field). The relationship between the vortex density and the magnetic induction is assumed to be local, which is permissible in bulk samples with dimensions much greater than the London penetration depth λ_L, if the magnetic induction changes little on scale lengths of order λ_L.[16–18] We also neglect viscous friction forces, which are proportional to the velocity of the vortices. This is permissible if the external magnetic field varies sufficiently slowly. Finally, we assume that the pinning force on each individual vortex is independent of the spatial coordinates and the magnetic induction. According to the conventional critical-state model, this approximation corresponds to neglecting the dependence of the critical current density on the magnetic induction. The theory is constructed for the case in which all physical quantities depend only on the spatial coordinate x, which corresponds to considering a half space or a sufficiently thick slab.

In Sec. 2 of this paper we formulate the basic equations of the kinetic theory. The first step is to justify the need to introduce two functions of the coordinates and angles. One of them is the distribution function of the vortices, $n(x,\vartheta)$, and the other, the velocity $V(x,\vartheta)$ at which they move. The first of the equations for these functions is a kinetic equation with a collision integral that takes the crossing of vortices, which changes their orientation, into account. The second equation describes force balance for the vortices. Collective effects associated with the crossing of the vortices are included in the magnetic force acting on the vortices.

Section 3 is devoted to an analysis of the electrodynamic consequences of the kinetic equations. It is shown that vortex crossing can lead to a substantial reduction (collapse) in the gradient of the absolute value of the magnetic induction. The transition from the kinetic equations to the equations of the two-velocity hydrodynamic model is made in this same section.

2. BASIC EQUATIONS OF THE MODEL

We consider a superconductor occupying a half space ($x>0$) or slab ($-d/2<x<d/2$) in an external magnetic field which varies both in magnitude and direction, but is always parallel to the surface of the sample. In this case, vortices which enter the superconductor differ in direction from vortices which are already present in the sample. As a result, a situation arises in which a given physically small superconducting volume may contain vortices with different angles ϑ of inclination to some given z axis. The variation of the external magnetic field $\mathbf{H}_0(t)$ causes the vortices to move at a velocity which may depend not only on the coordinate x, but also on the angle ϑ. In the following, we describe the state and motion of the vortical system with the aid of averaged quantities that apply to large groups of vortices with similar angles. To do this we introduce the vortex number density $n(x,\vartheta)$ and velocity density $V(x,\vartheta)$, which depend on position x and angle ϑ. We define the vortex density in such a way that $n(x\vartheta)dx\,d\vartheta$ is the number of vortices with angles within the interval $(\vartheta,\vartheta+d\vartheta)$ in the spatial neighborhood $(x,x+dx)$. $V(x,\vartheta)$ is the average velocity of the vortices in the neighborhood of a point x with angles close to ϑ.

In the following we assume that the external magnetic field varies so slowly that the magnetic force on a vortex is balanced only by the pinning force f_p, while the viscous drag forces ηV, where η is the coefficient of viscosity, can be neglected. This means that the theory we are developing is valid at relatively low vortex velocities $V<f_p/\eta$. We assume also that the magnetic field nevertheless varies fast enough that it is possible to neglect flow creep, which leads to a slight logarithmic relaxation of the magnetic moment and a uniform distribution of the magnetic field within the superconductor for $H_0=\text{const}$. We assume that the characteristic time for changes in the magnetic field is much shorter than the characteristic relaxation time. Under these assumptions, the velocity $V(x,\vartheta)$ instantaneously vanishes at all points in the sample when the external magnetic field $\mathbf{H}_0$ ceases to vary. These approximations correspond to the conditions for applicability of the critical-state model.

Here we note the fundamental differences between the statement of this kinetic problem and the ordinary kinetic description of the behavior of a system of particles. Usually the kinetic equation is formulated for the distribution function of the particles with respect to their positions and momenta,whereas solving the kinetic equation makes it possible to find any kinetic characteristic of the system. In the situation we are considering, the interacting objects (vortices) have no kinetic energy. Thus, their velocity is entirely determined by the force balance condition, which has the same form for all vortices with given x and ϑ. Therefore, all vortices with given x and ϑ move at the same velocity $V(x,\vartheta)$, i.e., the velocity distribution of the vortices is a δ-function. Hence, the kinetic equation is formulated for a distribution function $n(x,\vartheta)$ in which only the position x and angle ϑ appear as an argument. In the theory, however, there is another function $V(x,\vartheta)$ to be determined. In order to find both unknown functions $n(x,\vartheta)$ and $V(x,\vartheta)$, besides the kinetic equation we have to formulate a second, independent, equation, which is the aforementioned condition for the balance of forces acting on a vortex. The present section of the paper is devoted to deriving this system of equations.

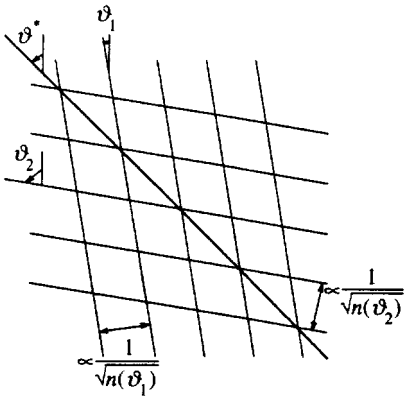

FIG. 1. Intersection of two series of vortices.

A. Kinetic equation

This section is devoted to deriving the kinetic equation for the vortex number density $n(x,\vartheta)$ as vortices move and intersect. Above all, it must be pointed out that vortices in a rigid superconductor cannot be made to rotate by an external magnetic field. In fact, the gain in magnetic energy upon rotating a vortex as a whole is proportional to its length, while the energy loss involved in overcoming the pinning force is proportional to the square of the length. Thus, this process is energetically forbidden for infinitely long vortices. This means that the orientation of the vortices can be changed only through their mutual crossing, followed by cutting and rotation of small segments of the new vortices relative to one another.[10] When this circumstance is taken into account, the kinetic equation can be written in the form

$$\frac{\partial n(x,\vartheta)}{\partial t}+\frac{\partial[n(x,\vartheta)V(x,\vartheta)]}{\partial x}=I(n), \tag{2}$$

where $I(n)=I_+(n)-I_-(n)$ is the collision integral and $I_+(n)$ and $I_-(n)$ are terms describing the ''arrival'' and ''departure'' of vortices in a state with angle near ϑ.

To derive the collision integral, we consider two intersecting series of vortices with angles ϑ_1 and ϑ_2 (see Fig. 1). If the crossing of these series is accompanied by reconnection, all the newly formed vortices will turn out to be oriented along a common direction ϑ^*. Given that the distances between the vortices in the series are proportional to $1/\sqrt{n(\vartheta_1)}$ and $1/\sqrt{n(\vartheta_2)}$, some simple geometrical calculations yield an equation for the direction of the newly formed vortices:

$$\vartheta^*(x,\vartheta_1,\vartheta_2)=\frac{\vartheta_1\sqrt{n(\vartheta_1)}+\vartheta_2\sqrt{n(\vartheta_2)}}{\sqrt{n(\vartheta_1)}+\sqrt{n(\vartheta_2)}}. \tag{3}$$

Here it is assumed that $|\vartheta_1-\vartheta_2|\ll 1$. The number of vortex trains per unit length along the x axis with angle ϑ_1, and crossing per unit time with trains that have angle ϑ_2, obeys the equation

$$\nu(x,\vartheta_1,\vartheta_2)=\sqrt{n(x,\vartheta_1)}\sqrt{n(x,\vartheta_2)}|V(x,\vartheta_1)-V(x,\vartheta_2)|, \tag{4}$$

where $V(x,\vartheta_1)$ and $V(x,\vartheta_2)$ are the average velocities of the vortices oriented along angles ϑ_1 and ϑ_2, respectively.

If the intersection of these series is accompanied by reconnection, then they go into a state with angle ϑ^*. We assume that the number of vortices per unit interval of x with angles ϑ^* increases by

$$N(x,\vartheta_1,\vartheta_2)=\sqrt{n(x,\vartheta_1)}+\sqrt{n(x,\vartheta_2)}. \tag{5}$$

Using these expressions, we write the incoming term of the collision integral in the form

$$I_+=\frac{1}{2}\int d\vartheta_1\, d\vartheta_2\, P(x,\vartheta_1,\vartheta_2)\nu(x,\vartheta_1,\vartheta_2) \times N(x,\vartheta_1,\vartheta_2)\delta[\vartheta-\vartheta^*(x,\vartheta_1,\vartheta_2)]. \tag{6}$$

Here $P(x,\vartheta_1,\vartheta_2)=P(x,\vartheta_2,\vartheta_1)$ is the probability function for cutting of two intersecting vortex series, which is symmetric with respect to interchange of the angles. It is easy to show that the ''departure'' of vortices from the state with angle ϑ is given by

$$I_-=\int d\vartheta_2\, P(x,\vartheta,\vartheta_2)\nu(\vartheta,\vartheta_2)\sqrt{n(x,\vartheta)}. \tag{7}$$

Given the angular symmetry of $P(x,\vartheta_1,\vartheta_2)$, we obtain the following expression for the vortex collision integral:

$$I=\int d\vartheta_1\, d\vartheta_2\, P(x,\vartheta_1,\vartheta_2) \times n(x,\vartheta_1)\sqrt{n(x,\vartheta_2)}|V(x,\vartheta_1)-V(x,\vartheta_2)| \times\left[\delta\left(\vartheta-\frac{\vartheta_1\sqrt{n(\vartheta_1)}+\vartheta_2\sqrt{n(\vartheta_2)}}{\sqrt{n(\vartheta_1)}+\sqrt{n(\vartheta_2)}}\right)-\delta(\vartheta-\vartheta_1)\right]. \tag{8}$$

Here we note two important properties of this collision integral: it preserves both the total number of vortices and their average angle of inclination to the z axis. In other words, the two integrals $\int d\vartheta\, I(\vartheta)$ and $\int d\vartheta\, \vartheta I(\vartheta)$ are equal to zero.

An even more general property of the collision integral (8) can be derived. For sufficiently smooth functions φ of the angle ϑ, the following equality holds (see Appendix A):

$$\int d\vartheta\, I(\vartheta)\varphi(\vartheta)=-\frac{1}{4}\left.\frac{\partial^2\varphi}{\partial\vartheta^2}\right|_{\vartheta=\overline{\vartheta}} \times\int d\vartheta_1\, d\vartheta_2\, P(x,\vartheta_1,\vartheta_2)|V(x,\vartheta_1) -V(x,\vartheta_2)|\frac{n(x,\vartheta_1)n(x,\vartheta_2)}{\sqrt{n(x,\vartheta_1)}+\sqrt{n(x,\vartheta_2)}} \times(\vartheta_1-\vartheta_2)^2, \tag{9}$$

where the derivative $\partial^2\varphi/\partial\vartheta^2$ is calculated at the point $\vartheta=\overline{\vartheta}(x)$, which defines the direction of the average angle,

$$\overline{n}(x)\,\overline{\vartheta}(x)=\int d\vartheta\, \vartheta n(x,\vartheta). \tag{10}$$

In the latter expression, the total vortex density $\overline{n}(x)$ is defined as

$$\overline{n(x)}=\int d\vartheta\; n(x,\vartheta). \tag{11}$$

Using these properties of the collision integral, we obtain quasihydrodynamic transport equations for the total vortex density $\overline{n(x)}$ and the average angle of inclination $\overline{\vartheta(x)}$ (see Appendix A):

$$\frac{\partial\overline{n}}{\partial t}+\frac{\partial(\overline{n}\overline{V})}{\partial x}=0, \tag{12}$$

and

$$\frac{\partial(\overline{n}\overline{\vartheta})}{\partial t}+\frac{\partial(\overline{n}\overline{V}\overline{\vartheta})}{\partial x}=0, \tag{13}$$

where $\overline{V(x)}$ and $\overline{V\vartheta}$ are defined by

$$\overline{V}=\frac{1}{\overline{n}}\int d\vartheta\; V(x,\vartheta)n(x,\vartheta), \tag{14}$$

and

$$\overline{V\vartheta}=\frac{1}{\overline{n}}\int d\vartheta V(x,\vartheta)\vartheta n(x,\vartheta). \tag{15}$$

B. Force balance equation

In order to finish formulating the kinetic problem, the equation obtained in the previous part must be supplemented by an equation for the balance of forces on the vortices, as well as by a formula relating the magnetic induction to the vortex density. In the local limit, the relationship between the magnetic induction vector **B** and the vortex density averaged over the angle ϑ can be written as

$$\mathbf{B}=\Phi_0\int d\vartheta\; n(x,\vartheta)\mathbf{e}(\vartheta)\approx\overline{n(x)}\mathbf{e}(\overline{\vartheta}), \tag{16}$$

where $\mathbf{e}(\vartheta)$ is the unit vector that defines the direction of a vortex with angle of inclination ϑ to the z axis.

The balance equation represents the equality of the magnetic force resulting from the interaction of vortices with one another and the pinning force acting on the vortex system. In order to calculate the magnetic force, it is first necessary to find the magnetic part of the Gibbs energy G_{em}. Using the well known expression for G_{em} (see Shmidt and Mkrtchyan,[19] for example), in the local limit we obtain

$$\begin{aligned}G_{em}&=\frac{1}{8\pi}\int dx(\mathbf{B}^2-2\mathbf{H}_0\cdot\mathbf{B})\\&=\frac{\Phi_0^2}{8\pi}\int dx\; dx'd\vartheta\; d\vartheta'[\mathbf{e}(\vartheta)\cdot\mathbf{e}(\vartheta')]\\&\quad\times n(x',\vartheta')n(x,\vartheta)-\frac{\mathbf{H}_0\Phi_0}{4\pi}\\&\quad\times\int dx\; d\vartheta\;\mathbf{e}(\vartheta)n(x,\vartheta).\end{aligned} \tag{17}$$

The magnetic force $f_m(x,\vartheta)$ on a vortex subsystem with angles ϑ at the point x is given by the variation of the Gibbs energy G_{em} when the vortices are displaced by $\delta u(x,\vartheta)$:

$$\delta G_{em}=\int dx\; d\vartheta\; n\;(x,\vartheta)f_m(x,\vartheta)\delta u(x,\vartheta). \tag{18}$$

Unfortunately, it is impossible to calculate the magnetic force directly from Eqs. (17) and (18), since the functional G_{em} depends on the vortex density $n(x,\vartheta)$. This means that it is necessary to find the relationship between the change in the vortex density $\delta n(x,\vartheta)$ and a small displacement $\delta u(x,\vartheta)$ of the vortices. This sort of relationship is easily found with the aid of the kinetic equation. Integrating Eq. (2) over the short time interval τ during which the displacement $\delta u(x,\vartheta)=\tau V(x,\vartheta)$ takes place, and retaining only the leading terms in the displacement, we obtain

$$\delta n(x,\vartheta)=-\frac{\partial[n(x,\vartheta)\delta u(x,\vartheta)]}{\partial x}+I\tau. \tag{19}$$

Now, calculating the change in the magnetic part of the Gibbs energy (17) for a small change in the vortex density and using Eqs. (18) and (19), we find an expression for the magnetic force (see Appendix B):

$$\begin{aligned}\frac{4\pi}{\Phi_0}f_m(x,\vartheta)&=\mathbf{e}(\vartheta)\cdot\frac{\partial\mathbf{B}}{\partial x}+\frac{1}{2}\mathbf{e}(\overline{\vartheta})\cdot(\mathbf{B}-\mathbf{H}_0)\\&\quad\times\int\frac{P(\vartheta,\vartheta')n(\vartheta')d\vartheta'}{\sqrt{n(\vartheta)}+\sqrt{n(\vartheta')}}\\&\quad\times(\vartheta-\vartheta')^2\mathrm{sign}[V(\vartheta)-V(\vartheta')].\end{aligned} \tag{20}$$

The first term in Eq. (20) is associated with the Lorentz force on a vortex. This is the term which transforms to the ordinary expression for the magnetic force (see Ref. 18, for example) when all the vortices are parallel to one another and to the external field. The second term results solely from the intersection of vortices, and vanishes if they are all parallel to one another, or their cutting probability $P(\vartheta_1,\vartheta_2)=0$. This term results from some rather intriquing collective effects. It is clear that any translational displacement of a given vortex causes intersections and rotation of neighboring vortices, as well as of itself. In other words, during a displacement, not only does the state of the given vortex change, so does that of its neighbors. As a result, an additional collective contribution, and therefore an additional collective force, shows up in the Gibbs energy. From this standpoint, the vortex system is not a gas of independently moving vortices, but a liquid with a very strong intervortex interaction.

The total magnetic force on a vortex in the critical state must be balanced by an effective pinning force. This latter force includes both drag on the vortices owing to their displacement and drag owing to the way in which pinning forces affect the straightening of the vortices after they have intersected. In order to calculate this effective drag force, we compute the work δG_p and δG_{cut} expended by the pinning forces in displacing and straightening the vortices, respectively, as they undergo a small displacement $\delta u(x,\vartheta)$. Since work $f_p|\delta u_i|=f_p\delta u_i\,\mathrm{sign}\,V_i$ is performed during displacement of the ith vortex, the work δG_p to displace all the vortices is

$$\delta G_p = \int dx\ d\vartheta\ n(x,\vartheta)\,\delta u(x,\vartheta) f_p\ \mathrm{sign}[V(x,\vartheta)]. \quad (21)$$

The factor sign(V) appears here because the work of overcoming the pinning forces is always positive, i.e, the pinning force is directed opposite the vortex velocity.

The second term in the drag on vortex motion is due to energy losses in straightening the vortices formed by cutting. In order to estimate this force, we assume that the vortices are perfectly elastic. This means that a new vortex formed after intersection, and consisting of segments along the directions of the vortices before they intersected, must be straightened. Then the segments of a given vortex turn toward one another, performing work against the pinning force (see Fig. 1). The work for a single vortex formed by the intersection of vortex trains with angles ϑ_1 and ϑ_2 is

$$A = \frac{1}{2} f_p \frac{a(\vartheta_1)a(\vartheta_2)}{a(\vartheta_1)+a(\vartheta_2)} \propto \frac{1}{2} f_p \frac{1}{\sqrt{n(x,\vartheta_1)}+\sqrt{n(x,\vartheta_2)}}, \quad (22)$$

where $a(\vartheta) \propto 1/\sqrt{n(\vartheta)}$ is the characteristic distance between vortices in a train with an angle of inclination ϑ. Using the latter equation and Eq. (4), which gives the number of intersecting vortices, we obtain

$$\delta G_{\mathrm{cut}} = \frac{1}{2}\int dx\ d\vartheta\ d\vartheta' f_p P(x,\vartheta,\vartheta') \times \sqrt{n(x,\vartheta)}\sqrt{n(x,\vartheta')}\,\delta u(x,\vartheta) \times \mathrm{sign}[V(x,\vartheta)-V(x,\vartheta')]. \quad (23)$$

The total drag force f_p^{eff} acting on a vortex is given by

$$\delta G_p + \delta G_{\mathrm{cut}} = \int dx\ d\vartheta\ f_p^{\mathrm{eff}} n(x,\vartheta)\,\delta u(x,\vartheta). \quad (24)$$

With the aid of Eqs. (21) and (23), this equation yields an effective drag force

$$f_p^{\mathrm{eff}} = f_p\ \mathrm{sign}\ V(x,\vartheta) + \frac{1}{2} f_p \int d\vartheta' P(x,\vartheta,\vartheta') \frac{\sqrt{n(x,\vartheta')}}{\sqrt{n(x,\vartheta)}} \times \mathrm{sign}[V(x,\vartheta)-V(x,\vartheta')]. \quad (25)$$

Using the condition that the sum of all forces on a vortex equals zero, we obtain:

$$\frac{\partial B}{\partial x} + (\vartheta - \overline{\vartheta}) B \frac{\partial \overline{\vartheta}}{\partial x} + \frac{1}{2}[B - H_0\ \cos(\overline{\vartheta} - \vartheta_0)] \times \int \frac{P(\vartheta,\vartheta')n(\vartheta')d\vartheta'}{\sqrt{n(\vartheta)}+\sqrt{n(\vartheta')}} (\vartheta-\vartheta')^2 \times \mathrm{sign}[V(\vartheta)-V(\vartheta')] = -\frac{4\pi}{c} J_c\ \mathrm{sign}^*[V(\vartheta)] - \frac{2\pi}{c} J_c \int d\vartheta' P(\vartheta,\vartheta') \frac{\sqrt{n(\vartheta')}}{\sqrt{n(\vartheta)}} \times \mathrm{sign}[V(\vartheta)-V(\vartheta')], \quad (26)$$

where $J_c = c f_p/\Phi_0$ is the critical current density, ϑ_0 is the angle of inclination of the external magnetic field **H** to the z axis, and sign(0)=0. The function sign*(x) introduced here coincides with sign(x) everywhere except $x=0$, where sign*(0) lies within the interval $(-1,1)$. It is necessary to introduce sign*[$V(\vartheta)$] because at a given point in space there can be both moving and motionless vortices, while for the latter the drag on their motion (the resting frictional force) is not generally equal to zero. This was also the reason for replacing sign[$V(x,\vartheta)$] by sign*[$V(x,\vartheta)$]. At the same time, in all cases in which the arguments of the sign function are relative velocities [the velocity difference $V(x,\vartheta)-V(x,\vartheta')$], the function sign[$V(x,\vartheta)-V(x,\vartheta')$] is retained. Indeed, in this case the vortices with angles ϑ and ϑ' do not intersect and the drag force corresponding to straightening is strictly equal to zero. If some of the vortices at any point move, while some are at rest, then this creates the impression that for the stationary vortices sign*(0) and, therefore, the right-hand side of Eq. (26), are undetermined. Since the velocity of these vortices is zero, however, the temporal and spatial variation in their density can be entirely determined from the kinetic equation, while in this case Eq. (26) serves for finding the value of sign*[$V(x,\vartheta)=0$]. The boundary angle between the region corresponding to $V(x,\vartheta)=0$ and the interval where $V(x,\vartheta)$ is nonzero for the distribution $n(x,\vartheta)$ can be determined from the continuity condition for V with respect to the variable ϑ.

The boundary conditions for Eqs. (2), (16), and (26) are that the magnetic induction at the sample boundary, $\mathbf{B}(0,t)$, equals the external field $\mathbf{H}_0(t)$, as well as the requirement that $V(x,t)$ be continuous everywhere that the vortex density is nonzero.

In the general case, the solution of the problem in the framework of the kinetic theory given here involves determining the vortex and velocity distributions $n(x,\vartheta)$ and $V(x,\vartheta)$ for a given quasistationary variation in the external magnetic field with time, $\mathbf{H}_0(t)$. To do this, Eqs. (2), (16), and (26) must be solved for the unknown functions. In the following sections we use Eqs. (2), (16), and (26) to establish the electrodynamics of a superconductor in an external magnetic field whose direction varies, and to analyze the most important magnetic properties of objects of this type.

3. ELECTRODYNAMIC EQUATIONS

The solution for the dynamics of a vortex lattice in terms of the proposed kinetic model is very complicated. To interpret electrodynamic experiments, however, there is usually no need to calculate the details of the vortex density distribution with respect to their angle of inclination to the z axis. Thus, it is appropriate to reduce the basic equations of the theory to a simpler system of electrodynamic equations which determine only the spatial distribution of the induction (its magnitude and angle of inclination to the z axis). This is the main purpose of the present section.

A. Suppression of the gradient of the magnitude of the magnetic induction

We first analyze in broad strokes the features of the distribution of the magnetic induction that follow from the force

balance Eq. (26). Multiplying Eq. (26) by the vortex density $n(x,\vartheta)$ and integrating the resulting equation with respect to ϑ, we obtain

$$\frac{\partial B}{\partial x} = -\frac{4\pi J_c}{cB}\int d\vartheta\ \Phi_0 n(\vartheta)\,\mathrm{sign}^*[V(\vartheta)]. \tag{27}$$

When the external magnetic field does not vary in direction and all the vortices are parallel to one another, the velocity V of the vortices has the same sign everywhere. Under these conditions Eq. (27) yields the usual Bean equation[1]: the vortex density $n=B/\Phi_0$ varies linearly in the depth of the superconductor, with slope $4\pi J_c/\Phi_0$.

In the other situation, when the vortices change direction in time, two fundamentally different cases may be realized. If essentially all vortices at a given point in space move in the same direction, then Eq. (27) transforms to an equation that is the same as the Bean equation,[1]

$$\frac{\partial B}{\partial x} = -\frac{4\pi}{c} J_c\ \mathrm{sign}(\overline{V}). \tag{28}$$

If, on the other hand, some of the vortices move in one direction and some in the other within a single spatial region, then the gradient of the induction is smaller. In fact, in this case, the values of $\mathrm{sign}^*(V)$ for two vortices moving in different directions are opposite. For those regions of the superconductor where the number of vortices moving in different directions are close to one another, there is essentially complete suppression of the gradient of the magnitude of the magnetic induction:

$$\partial B/\partial x \approx 0. \tag{29}$$

The collapse of the gradient of the vortex density corresponding to the transition from the hydrodynamic Eq. (28) to Eq. (29) must be taken into account when interpreting the large group of experiments that have been done in crossed magnetic fields.[3–5]

Equations (28) and (29) have a fairly simple form and clear physical significance. Unfortunately, they are not sufficient for solving actual electromagnetic problems. In fact, in order to determine which of the two equations must be used, it is first necessary to establish the velocity field $V(x,t)$, i.e., return to an analysis of the complicated original system of Eqs. (2), (16), and (26). For this reason there is some interest in constructing simpler models based on Eqs. (2), (16), and (26) that permit of a sufficiently simple and lucid physical interpretation.

B. Electrodynamics in the framework of a two-component model

Let us formulate a closed system of equations in terms of the simplest model assumption, in which the actual angular distribution of the vortices is replaced by

$$n(x,\vartheta) = \frac{1}{2} n(x)\{\delta[\vartheta-\vartheta_A(x)]+\delta[\vartheta-\vartheta_B(x)]\}. \tag{30}$$

In other words, we assume that at each point x, half of the vortices (type A vortices) have an angle ϑ_A and move at velocity V_A, while the rest are characterized by angle ϑ_B and velocity V_B. For concreteness, we assume that $\vartheta_A > \vartheta_B$.

The basic problem upon going to the two-component system is that in terms of this approximation it is impossible to satisfy the kinetic Eq. (2), since when vortices from groups A and B intersect, vortices develop which do not belong to either type.

Thus, instead of the kinetic equation, we make use of its hydrodynamic consequences: the transport equations for the total density (12) and the average angle (13). In the model of Eq. (30), these equations take the form

$$\frac{\partial n}{\partial t} + \frac{\partial[n(V_A+V_B)/2]}{\partial x} = 0, \tag{31}$$

and

$$\frac{\partial[n(\vartheta_A+\vartheta_B)]}{\partial t} + \frac{\partial[n(\vartheta_A V_A+\vartheta_B V_B)]}{\partial x} = 0. \tag{32}$$

We now analyze the force balance equation for two vortex sublattices A and B. The equation for component A has the form

$$\begin{aligned}\frac{\partial B}{\partial x} + \frac{(\vartheta_A-\vartheta_B)}{2} B\frac{\partial\overline{\vartheta}}{\partial x} + \frac{1}{2}[B-H_0\cos(\overline{\vartheta}-\vartheta_0)]\\ \times\frac{P\sqrt{n(x)}}{2\sqrt{2}}(\vartheta_A-\vartheta_B)^2\,\mathrm{sign}[V_A(x)-V_B(x)]\\ = -\frac{4\pi}{c}J_c\,\mathrm{sign}^*[V_A(x)] - \frac{2\pi}{c}J_c P\\ \times\mathrm{sign}[V_A(x)-V_B(x)],\end{aligned} \tag{33}$$

where the average angle $\overline{\vartheta}$ is given by $\overline{\vartheta}=(\vartheta_A+\vartheta_B)/2$. The force balance equation for group B can be obtained from Eq. (33) by replacing A with B.

In terms of the two component model, Eq. (27), which describes the suppression of the gradient of the magnetic induction, can be rewritten in the form

$$\frac{\partial B}{\partial x} = -\frac{2\pi}{c} J_c[\mathrm{sign}^*(V_A)+\mathrm{sign}^*(V_B)]. \tag{34}$$

Eliminating the gradient $\partial B/\partial x$ from the force balance Eq. (33) using Eq. (34), we obtain an equation for the spatial distribution of $\overline{\vartheta}$:

$$\begin{aligned}n(x)(\vartheta_A-\vartheta_B)B\frac{\partial\overline{\vartheta}}{\partial x} + [B-H_0\cos(\overline{\vartheta}-\vartheta_0)]P\sqrt{n(x)/8}\\ \times(\vartheta_A-\vartheta_B)^2\,\mathrm{sign}(V_A-V_B)\\ = -\frac{4\pi}{c}J_c[\mathrm{sign}^*(V_A)-\mathrm{sign}^*(V_B)]\\ -\frac{4\pi}{c}J_c P\,\mathrm{sign}(V_A-V_B).\end{aligned} \tag{35}$$

Unfortunately, in the two-component model, it is not possible to close the system of transport Eqs. (31) and (32) and the force balance equation for components A (33) and B without additional assumptions. In fact, there are five un-

known functions to be found, which we choose to be V_A, V_B, B, $\overline{\vartheta}$, and $\vartheta_A-\vartheta_B$. However, we have only four equations for these functions. In order to close this system, it is necessary to add one more phenomenological equation. As in Ref. 15, we assume that the characteristic scatter in the angles $\vartheta_A-\vartheta_B$ is proportional to the gradient of the average angle $(1/2)\partial(\vartheta_A+\vartheta_B)/\partial x$:

$$\vartheta_A-\vartheta_B\approx -l\ \mathrm{sign}(V_A-V_B)\frac{\partial\overline{\vartheta}}{\partial x}, \tag{36}$$

where l is a phenomenological parameter that determines the mean free path of the vortices between successive intersection processes in which old vortices are destroyed and new ones created. Qualitatively, Eq. (36) can be justified if we assume that vortices from some neighborhood of a given point x arrive at that point and carry different angles of inclination to the z axis with them (see Appendix C).

Using Eq. (36) it is possible to obtain a closed system of electrodynamic equations. Thus, for the case in which all vortices move in a single direction ($\mathrm{sign}\ V_A=\mathrm{sign}\ V_B$), the magnitude of the magnetic induction is calculated from Eq. (28), and the angular distribution is given by

$$\left\{B-\frac{l}{l^*}[B-H_0\ \cos(\overline{\vartheta}-\vartheta_0)]\right\}l\left(\frac{\partial\overline{\vartheta}}{\partial x}\right)^2=\frac{4\pi}{c}J_cP, \tag{37}$$

where l^* is defined by

$$l^*\approx\frac{2\sqrt{2}}{P\sqrt{n}}. \tag{38}$$

If, on the other hand, the vortices in groups A and B move in opposite directions ($\mathrm{sign}\ V_A=-\mathrm{sign}\ V_B$), then the distribution of B is found from Eq. (29) and the average angle is calculated with the aid of the equation

$$\left\{B-\frac{l}{l^*}[B-H_0\ \cos(\overline{\vartheta}-\vartheta_0)]\right\}l\left(\frac{\partial\overline{\vartheta}}{\partial x}\right)^2=\frac{4\pi}{c}J_c(P+2). \tag{39}$$

Besides these two cases, a situation can arise in which one of the vortex lattices is at rest within some region. In this region the magnitude of the induction gradient $\partial B/\partial x$ ranges from $4\pi J_c/c$ to zero, since the value of $\mathrm{sign}^*(0)$ ranges from -1 to 1. Eliminating the unknown parameter $\mathrm{sign}^*(0)$, we arrive at the equation

$$\frac{\partial B}{\partial x}-\frac{1}{2}\left\{B-\frac{l}{l^*}[B-H_0\ \cos(\overline{\vartheta}-\vartheta_0)]\right\}\times l\left(\frac{\partial\overline{\vartheta}}{\partial x}\right)^2\mathrm{sign}(V_A)=-\frac{4\pi}{c}J_c\ \mathrm{sign}(V_A)\left(1+\frac{P}{2}\right). \tag{40}$$

The boundaries between the regions in which the cases $\mathrm{sign}\ V_A=\mathrm{sign}\ V_B$, $\mathrm{sign}\ V_A=-\mathrm{sign}\ V_B$, and $V_B=0$ are realized can be determined from the transport Eqs. (31) and (32). The last of these equations can be rewritten in terms of V_A, V_B, n, and $\overline{\vartheta}$ as

$$\frac{\partial n\overline{\vartheta}}{\partial t}+\frac{1}{2}\frac{\partial n\overline{\vartheta}(V_A+V_B)}{\partial x}-\frac{1}{4}\frac{\partial}{\partial x}\left(n|V_A-V_B|l\frac{\partial\overline{\vartheta}}{\partial x}\right)=0. \tag{41}$$

Equations (28), (29), (31), (37), (39), (40), and (41) form a closed system of electrodynamic equations. If we set the length l equal to l^*, we arrive at the system of equations derived in Ref. 15.

In concluding this section we note that this two-component model is an extremely crude approximation. In fact, replacing the vortex density distribution function by just two vortex groups with angles ϑ_A and ϑ_B cannot be justified rigorously. Nevertheless, this model does take into account the fundamental kinetic features of a vortex lattice in crossed magnetic fields, so that the electrodynamic equations based on it can be used to interpret a variety of experimental results.

We now consider the limits of applicability of the two component model. All the basic equations of the theory and, therefore, of this model have been obtained assuming that the angular scatter of the vortices at each point of the sample is small, i.e., for $l(\partial\vartheta/\partial x)\ll 1$. Equation (37) implies that this condition is clearly violated in the neighborhood of points where $\cos(\overline{\vartheta}-\vartheta_0)=B(1-l/l^*)/H_0$. This limits the validity of the two-component model.

CONCLUSION

In this paper we have examined the behavior of a vortex system in a superconductor located in a magnetic field that varies in a quasistationary fashion, both in magnitude and in direction. In this case the vortices have different spatial orientations. In the proposed kinetic theory, the behavior of the vortex system is described by the kinetic Eq. (2) and the force balance Eq. (26), which determine the spatial distribution of the vortex density $n(x,\vartheta)$ and velocity $V(x,\vartheta)$. The collision integral (8) in the kinetic Eq. (2) describes the change in the vortex density $n(x,\vartheta)$ owing to their mutual intersections accompanied by a change in orientation. It has been shown that the collision integral conserves the total number of vortices and their average angle of inclination.

Besides the direct interactions of vortices with one another, an external magnetic field, and pinning centers, the force balance equation includes the collective interaction associated with the intersection and scattering of vortices on one another. This kind of collective vortex interaction makes an additional contribution to the magnetic force, which can cause the vortices to attract one another. This, in turn, produces counterpropagating motion of the vortices which suppresses the gradient of the magnitude of the magnetic induction in a superconductor.

We have examined the transition to a simpler two component model in which the actual distribution of the vortices with respect to their angles of inclination is replaced by a distribution in which the vortices have only two angles of inclination at every point, $\vartheta_A(x)$ and $\vartheta_B(x)$. In order to obtain a closed system of equations in the model, an additional phenomenological condition is added to couple the angular scatter of the vortices to the gradient of the average angle and to the mean free path. The resulting two-component hy-

drodynamic model makes it possible to interpret the large number of experiments on the electrodynamics of superconductors in crossed magnetic fields (see Ref. 15 and the literature cited therein).

This work was performed under the auspices of the State Program on Superconductivity (Projects 96046 and 95046), and partially supported by the Russian Fund for Fundamental Research (Project 97-02-16399).

APPENDIX A

The purpose of this appendix is to derive Eq. (9), along with the transport Eqs. (12) and (13). Multiplying Eq. (8) by $\varphi(\vartheta)$ and integrating, we find

$$J=\int d\vartheta_1\, d\vartheta_2\, P(\vartheta_1,\vartheta_2) \times n(x,\vartheta_1)\sqrt{n(x,\vartheta_2)}\,|V(x,\vartheta_1)-V(x,\vartheta_2)| \times\left[\varphi\left(\frac{\vartheta_1\sqrt{n(\vartheta_1)}+\vartheta_2\sqrt{n(\vartheta_2)}}{\sqrt{n(\vartheta_1)}+\sqrt{n(\vartheta_2)}}\right)-\varphi(\vartheta_1)\right]. \tag{A1}$$

In the following we assume that the scatter in the vortex angles about the average value $\overline{\vartheta}$ at a given point in space is small. In this case,

$$\varphi\left(\frac{\vartheta_1\sqrt{n(\vartheta_1)}+\vartheta_2\sqrt{n(\vartheta_2)}}{\sqrt{n(\vartheta_1)}+\sqrt{n(\vartheta_2)}}\right) \approx\varphi(\overline{\vartheta})+\left.\frac{\partial\varphi}{\partial\vartheta}\right|_{\overline{\vartheta}}\left(\frac{\vartheta_1\sqrt{n(\vartheta_1)}+\vartheta_2\sqrt{n(\vartheta_2)}}{\sqrt{n(\vartheta_1)}+\sqrt{n(\vartheta_2)}}-\overline{\vartheta}\right) +\frac{1}{2}\left.\frac{\partial^2\varphi}{\partial\vartheta^2}\right|_{\overline{\vartheta}}\left(\frac{\vartheta_1\sqrt{n(\vartheta_1)}+\vartheta_2\sqrt{n(\vartheta_2)}}{\sqrt{n(\vartheta_1)}+\sqrt{n(\vartheta_2)}}-\overline{\vartheta}\right)^2, \tag{A2}$$

and

$$\varphi(\vartheta_1)=\varphi(\overline{\vartheta})+\left.\frac{\partial\varphi}{\partial\vartheta}\right|_{\overline{\vartheta}}(\vartheta_1-\overline{\vartheta})+\frac{1}{2}\left.\frac{\partial^2\varphi}{\partial\vartheta^2}\right|_{\overline{\vartheta}}(\vartheta_1-\overline{\vartheta})^2. \tag{A3}$$

Substituting these expansions of the function φ into Eq. (A1), we obtain

$$J=J^{(1)}+J^{(2)}, \tag{A4}$$

where

$$J^{(1)}=\left(\left.\frac{\partial\varphi}{\partial\vartheta}\right|_{\overline{\vartheta}}-\overline{\vartheta}\left.\frac{\partial^2\varphi}{\partial\vartheta^2}\right|_{\overline{\vartheta}}\right)\int d\vartheta_1\, d\vartheta_2\, P(\vartheta_1,\vartheta_2) \times\frac{n(\vartheta_1)n(\vartheta_2)}{\sqrt{n(\vartheta_1)}+\sqrt{n(\vartheta_2)}}|V(\vartheta_1)-V(\vartheta_2)|(\vartheta_2-\vartheta_1), \tag{A5}$$

and

$$J^{(2)}=\frac{1}{2}\left.\frac{\partial^2\varphi}{\partial\vartheta^2}\right|_{\overline{\vartheta}}\int d\vartheta_1\, d\vartheta_2\,\frac{P(\vartheta_1,\vartheta_2)n(\vartheta_1)n(\vartheta_2)}{[\sqrt{n(\vartheta_1)}+\sqrt{n(\vartheta_2)}]^2} \times|V(\vartheta_1)-V(\vartheta_2)|(\vartheta_2-\vartheta_1)[2\sqrt{n(\vartheta_1)}\vartheta_1 +\sqrt{n(\vartheta_2)}\vartheta_2+\sqrt{n(\vartheta_2)}\vartheta_1]. \tag{A6}$$

The integral $J^{(1)}$ is equal to zero. Indeed, making the substitution $\vartheta_1\to\vartheta_2$, $\vartheta_2\to\vartheta_1$ in Eq. (A5), we arrive at the equation $J^{(1)}=-J^{(1)}$, which implies that $J^{(1)}=0$. Making the same change of variables in Eq. (A6), we obtain

$$J^{(2)}=\frac{1}{2}\left.\frac{\partial^2\varphi}{\partial\vartheta^2}\right|_{\overline{\vartheta}}\int d\vartheta_1\, d\vartheta_2\,\frac{P(\vartheta_1,\vartheta_2)n(\vartheta_1)n(\vartheta_2)}{[\sqrt{n(\vartheta_1)}+\sqrt{n(\vartheta_2)}]^2} \times|V(\vartheta_1)-V(\vartheta_2)|(\vartheta_1-\vartheta_2) \times[2\sqrt{n(\vartheta_2)}\vartheta_2+\sqrt{n(\vartheta_1)}\vartheta_1+\sqrt{n(\vartheta_1)}\vartheta_2]. \tag{A7}$$

Therefore, the desired integral J can be written in the form of an average of Eqs. (A6) and (A7). After some simple algebraic transformations, we arrive at Eq. (9).

Setting $\varphi=1$ and $\varphi=\vartheta$ in Eq. (9), we find that the collision integral (8) conserves the total number of vortices and the average angle, i.e., the following equations hold:

$$\int d\vartheta\; I(\vartheta)=0, \quad \int d\vartheta\;\vartheta I(\vartheta)=0. \tag{A8}$$

Integrating Eq. (2) with respect to ϑ and using Eq. (A8), we find the first transport equation (12). In order to obtain the second transport equation (13), it is sufficient to multiply the kinetic equation (2) by ϑ and integrate, using Eq. (A8). Thus, property (9) of the collision integral and the transport equations (12) and (13) have been proven.

APPENDIX B

The purpose of this appendix is to derive an expression for the magnetic force on the vortices, taking their intersections into account. A displacement $\delta u(x,\vartheta)$ of the vortices causes their distribution to change by a small amount $\delta n(x,\vartheta)$. This leads to a change δG_{em} in the magnetic energy of the system, which, given Eq. (17), can be written in the form

$$\delta G_{em}=\frac{\Phi_0}{4\pi}\int dx\, d\vartheta\;\mathbf{e}(\vartheta)\cdot[\mathbf{B}(x)-\mathbf{H}_0]\delta n(x,\vartheta). \tag{B1}$$

Substituting Eq. (19) into this expression and integrating by parts, we obtain

$$\delta G_{em}=\frac{\Phi_0}{4\pi}\int dx\, d\vartheta\; n(x,\vartheta)\delta u(x,\vartheta)\mathbf{e}(\vartheta)\cdot\frac{\partial\mathbf{B}}{\partial x} +\frac{\Phi_0}{4\pi}\int dx\, d\vartheta\;\delta I(n)\mathbf{e}(\vartheta)\cdot[\mathbf{B}(x)-\mathbf{H}_0], \tag{B2}$$

where $\delta I=I\tau$. On the other hand, using Eq. (9), the equation $\partial^2\mathbf{e}/\partial\vartheta^2=-\mathbf{e}$, and the symmetry of the probability function $P(\vartheta,\vartheta')$, we find

$$\int d\vartheta\; I(n)\tau\mathbf{e}(\vartheta) =\frac{1}{2}\int d\vartheta\;\mathbf{e}(\overline{\vartheta})n(x,\vartheta)\delta u(x,\vartheta)$$

$$\times \int \frac{P(\vartheta,\vartheta')n(\vartheta')d\vartheta'}{\sqrt{n(\vartheta)}+\sqrt{n(\vartheta')}}$$

$$\times \mathrm{sign}[V(\vartheta)-V(\vartheta')](\vartheta-\vartheta')^2. \quad \text{(B3)}$$

Substituting the latter expression into Eq. (B2), we obtain

$$\delta G_{em}=\frac{\Phi_0}{4\pi}\int dx\, d\vartheta\, n(x,\vartheta)\delta u(x,\vartheta)$$

$$\times\left\{ \mathbf{e}(\vartheta)\cdot\frac{\partial \mathbf{B}}{\partial x}+\frac{1}{2}\mathbf{e}(\overline{\vartheta})\cdot[\mathbf{B}(x)-\mathbf{H}_0]\right.$$

$$\times \int \frac{P(\vartheta,\vartheta')n(\vartheta')d\vartheta'}{\sqrt{n(\vartheta)}+\sqrt{n(\vartheta')}}$$

$$\left.\times(\vartheta-\vartheta')^2\ \mathrm{sign}[V(\vartheta)-V(\vartheta_1)]\right\}. \quad \text{(B4)}$$

Equation (20) for the magnetic force follows directly from this equation and Eq. (18).

APPENDIX C

This part of the paper is devoted to a qualitative justification of Eq. (36), as well as to estimating the mean free path l which appears in that equation.

Let us consider two vortices, denoted by 1 and 2, which formed at time $t-\tau$ at points x_1 and x_2 and then intersected at point x at time t. We calculate the angles of inclination ϑ_1 and ϑ_2 of these vortices to the z axis. Since vortices 1 and 2 appear as a result of the intersection of vortices of types A and B at points x_1 and x_2, Eq. (3) yields the following equations for the angles of inclination:

$$\vartheta_1=[\vartheta_A(x_1,t-\tau)+\vartheta_B(x_1,t-\tau)]/2=\overline{\vartheta}(x_1,t-\tau), \quad \text{(C1)}$$

and

$$\vartheta_2=[\vartheta_A(x_2,t-\tau)+\vartheta_B(x_2,t-\tau)]/2=\overline{\vartheta}(x_2,t-\tau). \quad \text{(C2)}$$

On reaching the point x, the vortices 1 and 2 become vortices of type A and B for this point and time t. As a result, we obtain an estimate for the difference in the angles of inclination of the type A and B vortices:

$$\vartheta_A-\vartheta_B=\overline{\vartheta}(x_1,t-\tau)-\overline{\vartheta}(x_2,t-\tau). \quad \text{(C3)}$$

Retaining only the linear terms in the expansion of the difference $\vartheta_A-\vartheta_B$ with respect to τ, i.e., $x-x_1$ and $x-x_2$, we obtain

$$\vartheta_A-\vartheta_B=\frac{\partial\overline{\vartheta}}{\partial x}(x_1-x_2). \quad \text{(C4)}$$

In order to determine the distance x_2-x_1 covered by a vortex without intersections, we pass from the laboratory reference frame K to a system K' attached to a type A vortex. In this system, the velocity of a B vortex is V_B-V_A. Over a time τ, $\sqrt{n/2}|V_A-V_B|\tau$ type B vortices intersect a type A vortex that is at rest. Since only every $1/P$-th intersection involves the creation of new vortices, as an estimate of the average time between intersections we obtain

$$\tau=\frac{\sqrt{2}}{P\sqrt{n}|V_A-V_B|}. \quad \text{(C5)}$$

In time τ in system K, an A vortex with velocity V_A moves a distance $V_A\tau=x-x_1$, and a B vortex with velocity V_B moves a distance $V_B\tau=x-x_2$. Thus, we obtain $x_2-x_1=(V_B-V_A)\tau$. Using this expression and Eqs. (C4) and (C5), we obtain the following estimate for the difference in the angles of inclination of the type A and B vortices:

$$\vartheta_A-\vartheta_B=-\mathrm{sign}(V_A-V_B)\frac{\sqrt{2}}{P\sqrt{n}}\frac{\partial\overline{\vartheta}}{\partial x}. \quad \text{(C6)}$$

Therefore, the length l can be estimated as

$$l=\frac{\sqrt{2}}{P\sqrt{n}}. \quad \text{(C7)}$$

[1] C. P. Bean, Phys. Rev. Lett. **8**, 250 (1962).
[2] I. V. Baltaga, L. M. Fisher, N. V. Il'in *et al.*, Phys. Lett. A **148**, 213 (1990).
[3] I. V. Baltaga, N. M. Makarov, V. A. Yampol'skii *et al.*, Fiz. Nizk. Temp. **21**, 411 (1995) [Low Temp. Phys. **21**, 320 (1995)].
[4] L. M. Fisher, A. V. Kalinov, I. F. Voloshin *et al.*, Solid State Commun. **97**, 833 (1996).
[5] S. K. Hasanain, S. Manzoor, and A. Amirabadizadeh, Supercond. Sci. Technol. **8**, 519 (1995).
[6] J. R. Clem, Phys. Rev. B **26**, 2463 (1982).
[7] J. R. Clem and A. Perez-Gonzalez, Phys. Rev. B **30**, 5041 (1984).
[8] R. Cave and M. A. R. LeBlanc, J. Appl. Phys. **53**, 1631 (1982).
[9] J. R. Clem, J. Appl. Phys. **50**, 3518 (1979).
[10] E. H. Brandt, J. R. Clem, and D. G. Walmsley, J. Low Temp. Phys. **37**, 43 (1979).
[11] E. H. Brandt, J. Low Temp. Phys. **39**, 41 (1980).
[12] R. Boyer, G. Fillion, and M. A. R. LeBlanc, J. Appl. Phys. **51**, 1692 (1980).
[13] L. M. Fisher, N. V. Il'in, I. F. Voloshin *et al.*, Physica C **197**, 161 (1992).
[14] I. V. Baltaga, K. V. Il'enko, N. M. Makarov *et al.*, Solid State Commun. **93**, 697 (1995).
[15] I. F. Voloshin, A. V. Kalinov, S. E. Savel'ev *et al.*, Zh. Éksp. Teor. Fiz. **111**, 1071 (1997) [JETP **84**, 592 (1997)].
[16] I. F. Voloshin, V. S. Gorbachev, S. E. Savel'ev *et al.*, JETP Lett. **59**, 55 (1994).
[17] L. M. Fisher, V. S. Gorbachev, S. E. Savel'ev *et al.*, Physica C **245**, 231 (1995).
[18] V. S. Gorbachev and S. E. Savel'ev, Zh. Éksp. Teor. Fiz. **107**, 1247 (1995) [JETP **80**, 694 (1995)].
[19] V. V. Shmidt and G. S. Mkrtchyan, Usp. Fiz. Nauk **112**, 459 (1974) [Sov. Phys. Usp. **17**, 170 (1974)].

Translated by D. H. McNeill

On the limiting velocity and forced motion of ferromagnetic domain walls in an external field perpendicular to the easy-magnetization axis

B. A. Ivanov

Insititute of Magnetism, Ukraine National Academy of Sciences, 252142 Kiev, Ukraine

N. E. Kulagin

State Institute of Problems in Radio Electronics, 103416 Moscow, Zelenograd, Russia

(Submitted 28 December 1996)
Zh. Éksp. Teor. Fiz. **112**, 953–974 (September 1997)

A theory is constructed for the dynamics and braking of domain walls in ferromagnets when a magnetic field is applied perpendicular to the axis of easy magnetization (i.e., a transverse field $H_\perp$). The theory is valid for velocities v up to the limiting domain wall velocity v_c. The Landau–Lifshitz equations in the dissipationless approximation are used to investigate the motion of domain walls and the change in the character of the wall motion as its velocity v approaches v_c. The force acting on a domain wall due to viscous friction is calculated within the framework of generalized relaxation theory, and the dependence of the domain wall velocity v on the forcing field H_z is investigated. Calculations of the braking force show that the contributions of various dissipation mechanisms to the friction force have different dependences on the domain wall velocity, which affects the form of the function $v=v(H_z)$. The shapes of the curves $v(H_z)$ differ very markedly from one another for different values of the field $H_\perp$. The theory developed here can be used to describe the experimental results, in particular the almost linear behavior of $v=v(H_z)$ for small $H_\perp$ and its strongly nonlinear behavior when $H_\perp \sim H_a$, whereas these data cannot be reconciled within the standard theory based on relaxation terms of Hilbert type. © *1997 American Institute of Physics.* [S1063-7761(97)01409-1]

1. INTRODUCTION

The motion of domain walls in ferromagnets is of considerable interest both to experimentalists and theorists as an example of strongly nonlinear wave dynamics realized in the magnetization of the material (magnetic solitons). It is also important for practical applications (see Refs. 1–4). From an experimental point of view, two questions are of interest: what is the limiting velocity of a domain wall, and how can we study the dependence of the domain-wall velocity v on the projection of the field onto the axis of easy magnetization (the forcing field H_z)?

The structure of a planar domain wall moving with a velocity v that is not small is given by the soliton solution to the Landau–Lifshitz equations without dissipation, in the form $\mathbf{M}=\mathbf{M}(\xi)$, $\xi=y-vt$, and $\mathbf{M}$ is the magnetization vector of the ferromagnet. The domain-wall velocity v, which enters into the equation for $\mathbf{M}(\xi)$ as a parameter, has a bifurcation value v_c such that solutions of domain wall type do not exist for $v>v_c$. This quantity therefore has the sense of a limiting velocity for domain walls. Soliton solutions are known only for the simplest Walker model (see Refs. 1–4 and below).

The braking of a domain wall, which determines the nature of forced motions of the domain wall under the action of a forcing field H_z, arises from relaxation terms in the Landau–Lifshitz equation, the analysis of whose structure has only begun (the use of ordinary relaxation terms of the Hilbert form leads to a number of inconsistencies with experiment; see the discussion of this question in Refs. 1 and 2). In order to calculate the friction force it is also necessary to know the form of the function $\mathbf{M}(\xi)$.

It is especially important to investigate the effect of an external magnetic field perpendicular to the axis of easy magnetization of the ferromagnet (the transverse field $\mathbf{H}_\perp$). Above all, this stems from the observation that the value of v_c is increased considerably even at rather small values of this field in the range $4\pi M_0 \ll H_\perp \ll H_a$, where $H_\perp=|\mathbf{H}_\perp|$, $M_0=|\mathbf{M}|$ is the saturation magnetization, and H_a is the anisotropy field, as pointed out in Ref. 2. (In what follows, we only discuss the case $q=H_a/4\pi M_0 \gg 1$, which applies, in particular, to magnetic bubble materials; q is the figure of merit of these materials.) In particular, the authors of Ref. 5, who investigated the dynamics of domain walls with regard to the problem of increasing the speed of response of magnetooptic light modulators, noted a sharp increase in domain wall velocity in the presence of a transverse field $\mathbf{H}_\perp$. An increase in the limiting velocity of domain walls to a value of order several km/s for $H_\perp$ of order $0.7H_a$ was observed by the authors of Ref. 6. Existing theories of domain wall motion in the presence of a transverse field are valid only when $H_\perp \ll H_a$ (see Ref. 2) or for $H_\perp \to H_a$ (see Refs. 7–9). In Ref. 9 we pointed out that the limiting velocity is governed by a variety of mechanisms at large and small values of the field. In particular, as the forcing field increases, the velocity v_c increases when $H_\perp \ll H_a$ and decreases when $H_\perp \approx H_a$. The braking of domain walls was investigated in the low-velocity limit in Refs. 8–12. The authors of Ref. 12 reported qualitative agreement between their experimental measure-

ments of how domain wall motion is affected by $H_\perp$ and the calculations of Refs. 8–10.

The theory we construct in this paper describes the dynamics and braking of a domain wall for arbitrary values of the transverse field and for velocities v up to v_c. In Sec. 2 we describe our model, and then investigate the motion of a domain wall in the dissipationless approximation. We show that as the field increases, the character of the wall behavior changes as its velocity v approaches v_c. In Sec. 3 we construct approximate solutions that enable us to describe the dynamics of a domain wall for fields that are not too large, actually up to $0.6H_a$. In Sec. 4 we investigate numerically and analytically the case of strong fields. In Sec. 5, based on the distribution of magnetization in a moving domain wall obtained above within the framework of generalized relaxation theory, we calculate the force of viscous friction acting on the wall. Calculation of the braking force shows that the contribution of various dissipation mechanisms to the frictional force have different dependences on the domain wall velocity, which affects the form of the function $v=v(H_z)$. The shapes of the curves $v(H_z)$ differ very markedly from one another for different values of field $H_\perp$. In the concluding section we show that our theory enables us to describe the results of experiment,[6] in particular the almost linear behavior of $v=v(H_z)$ for small $H_\perp$, and its strong linearity for $H_\perp \sim H_a$.

2. MODEL

Let us write the Landau–Lifshitz equation for the magnetization vector of a ferromagnet $\mathbf{M}$ in the form

$$\frac{\partial \mathbf{M}}{\partial t} = -g(\mathbf{M}\times\mathbf{F}) + \mathbf{R}, \tag{1}$$

where g is the gyromagnetic ratio, $\mathbf{F}=-\delta W/\delta \mathbf{M}$ is the effective field of the ferromagnet, and $W=W\{\mathbf{M}\}$ is its energy written in the form of a functional of the magnetization vector; see Refs. 1–4. $\mathbf{R}$ is a relaxation term, whose structure we will discuss below in Sec. 5.

We begin with an expression of the form

$$W\{\mathbf{M}\} = \int d\mathbf{r}\left\{\frac{\alpha}{2}(\nabla\cdot\mathbf{M})^2 + w(\mathbf{M})\right\}, \tag{2}$$

where α is the inhomogeneous exchange constant, $w(\mathbf{M})$ is the energy of the relativistic interaction, including the anisotropy energy, the external field, and the demagnetizing field. If the dissipation is weak, i.e., the corresponding dimensionless relaxation constant is small, we can postulate a structure for the domain wall based on the dissipationless approximation (for $\mathbf{R}=0$). In this case it is convenient to start with the equations for the unit vector $\mathbf{m}=\mathbf{M}/M_0$, which in angular coordinates are

$$m_z=\cos\theta, \quad m_x=\sin\theta\cos\varphi, \quad m_y=\sin\theta\sin\varphi. \tag{3}$$

The equation for the variables θ and φ can be written in the usual form:

$$gM_0\sin\theta\frac{\partial\theta}{\partial t} = \alpha\nabla^2\theta - \alpha(\nabla\theta)^2\sin\theta\cos\theta - \frac{\partial w}{\partial\theta},$$

$$gM_0\sin\theta\frac{\partial\varphi}{\partial t} = -\alpha\nabla(\sin^2\theta\nabla\varphi) + \frac{\partial w}{\partial\varphi}.$$

Let us write the relativistic interaction energy, including the energy in the external transverse magnetic field, in the form

$$w = \frac{M_0^2}{2}\beta\sin^2\theta\,(1+\rho\sin^2\varphi) - h\beta M_0^2\sin\theta\cos\varphi, \tag{4}$$

which is typical of a rhombic ferromagnet with axes x,y,z along the principal axes; β and $\rho\beta$ are anisotropy constants. Let $\beta>0$, $\rho>0$; then z is the axis of easy magnetization of the ferromagnet and y is the hardest axis. We assume that the transverse field $\mathbf{H}_\perp$ is directed along the intermediate axis x, $h=H_\perp/H_a$, $H_a=\beta M_0$. Note that this expression is also used for uniaxial ferromagnets, taking into account the energy of the demagnetizing field in the local (Vinterovskii) approximation. Then $q=1/\rho$ (see Refs. 2 and 4). Since our problem is one-dimensional, the Vinterovskii approximation is a good one. In what follows we sometimes neglect the anisotropy in the basal plane, assuming that ρ is not only considerably less than unity, but is much less than h.

The equations for $\varphi=\varphi(\xi)$, $\theta=\theta(\xi)$, taking into account the explicit form of the energy w Eq. (4), are easily written

$$\Delta^2[\theta'' - (\theta')^2\sin\theta\cos\theta] - \sin\theta\cos\theta\,(1+\rho\sin^2\varphi) + h\cos\theta\cos\varphi = \frac{v}{c}\varphi'\sin\theta, \tag{5}$$

$$\Delta^2(\varphi'\sin^2\theta)' - \sin\theta\sin\varphi\,(h+\rho\sin\theta\cos\varphi) = -\frac{v}{c}\theta'\sin\theta. \tag{6}$$

Here we introduce characteristic values of the domain wall thickness $\Delta=\sqrt{\alpha/\beta}$ and velocity $c=gM_0\sqrt{\alpha\beta}$. The quantity c is of the same order of magnitude as the magnon phase velocity $v_{(+)}$ (see below), and coincides with $v_{(+)}$ and the limiting velocity of the domain wall v_c as $h\to 1$; primes denote derivatives with respect to ξ, $\xi=y-vt$.

One important characteristic of a moving domain wall is the dependence of its energy on velocity. Using Eq. (8) (see below), we can write the expression for the domain wall energy per unit area σ in the form

$$\sigma = 2M_0^2\sqrt{\alpha\beta}\,I(v), \quad I(v)=\Delta\int_0^{+\infty} d\xi\,[(\theta')^2 + (\varphi')^2\sin^2\theta]. \tag{7}$$

The form of the integral $I(v)$ (like the integrals $\eta_\alpha(v)$ that describe the coefficients of nonlinear friction; see below) is determined by the characteristic distribution of magnetization in the domain wall.

The system of Eqs. (5) and (6) corresponds to a dynamical system with two degrees of freedom. In order for this

system to be integrable we must have two independent first integrals. In the general case of an arbitrary ferromagnet, there is a first integral of the form

$$\Delta^2[(\theta')^2+(\varphi')^2\ \sin^2\ \theta]-w(\theta,\varphi)=\text{const}. \tag{8}$$

However, it is not possible to construct another first integral, i.e., the problem is not integrable in general. Note that a number of examples of integrable equations similar to (5) and (6) were constructed by Eleonskiĭ and Kulagin in Ref. 13a for a motionless wall and various forms of anisotropy energy. However, even without the ''gyroscopic'' term in these equations (proportional to the velocity v), the system is not integrable when a transverse field is present. For $v\neq 0$, however, a single example is known for which a system like (5) and (6) is integrable—the case $h=0$ (the additional first integral other than Eq. (8) was derived in Ref. 13b for $h=0$). As Walker showed (see the review Refs. 1, 3, 4), when $h=0$ the Landau–Lifshitz equations (5) and (6) have an exact solution of the form $\varphi=\varphi_0=\text{const}$, $\theta=\theta(\xi)$, where

$$\cos\ \theta_0\equiv\pm\tanh[\xi/\Delta(\varphi_0)],$$

$$\Delta(\varphi_0)\equiv\Delta(1+\rho\ \sin^2\ \varphi_0)^{-1/2}. \tag{9}$$

The value of φ_0 is a function of the velocity of the domain wall. This function, which is given by Eq. (10) (see below), has a solution only for $v<v_W$, where v_W, referred to as the Walker limiting velocity (the Walker limit), is given by $v_W\equiv c(\sqrt{1+\rho}-1)$. When $\rho\ll 1$ we have $v_W\to g M_0(\rho/2)\sqrt{\alpha\beta}$, and v_W is small compared to the characteristic velocity $c\equiv g M_0\sqrt{\alpha\beta}$ we introduced above. The quantity v_W reduces to zero in the limit $\rho\to 0$. When $\rho\to 0$ we can neglect the change in the domain wall thickness and assume that the quantity $\Delta=\sqrt{\alpha/\beta}$.

As was noted previously,[8,9] when $\rho=0$, solutions can be constructed in two limits: large values of the field ($H_a-H_\perp\ll H_a$) and small values ($H_\perp\ll H_a$). We also can construct an analytic solution to (5) and (6) that includes the anisotropy in the basal plane when $H_\perp\ll\rho H_a$, based on perturbation theory in the form of an expansion in the parameter h/ρ. However, we will not discuss this approach here, since we can give a rather complete analysis of the more general case h, $\rho\ll 1$, including $h\sim\rho$. We will also show that the asymptotic solution obtained as $h\to 1$ is applicable only for a narrow range of field values. Numerical analysis of the problem over a wide range of values of h enables us to present a general picture of the behavior of domain walls in arbitrary fields. In the next two sections we begin with a discussion of asymptotic behavior for large and small transverse fields $H_\perp$, analyzing numerical data and using it to verify our approximations.

3. NONDISSIPATIVE DYNAMICS OF DOMAIN WALLS. SMALL FIELDS

For small fields we can exploit the similarity of our problem to that of Walker, and assume that when $\rho\ll 1$, $H_\perp\ll H_a$ the value of φ is not small, but varies slowly in space. This follows from a linearization of the problem with respect to φ and v, but can also be verified directly with the help of Eq. (6). Analysis of Eqs. (5) and (6) leads us to conclude that if the angle $\theta(\xi)$ varies over a distance of order $\Delta\xi\sim\Delta$, then the angle $\varphi(\xi)$ varies over a much larger distance $\Delta\xi\sim\Delta\sqrt{\beta/h}$. Therefore, within the domain wall (for $\xi\leqslant\Delta$), we may assume that $\varphi(\xi)\approx\varphi_0=\text{const}$ and is independent of ξ. The value of φ_0 can be found by integrating Eq. (6) with respect to ξ from $-\infty$ to ∞. Taking Eq. (7) into account, and the fact that for h, $\rho\ll 1$ the quantity $\varphi'\ll\theta'$ and $v_c\ll c$, it is easy to obtain the relation

$$\frac{v}{c}=\left[\rho\ \cos\ \varphi_0+\frac{\pi h}{2}\right]\sin\ \varphi_0. \tag{10}$$

From Eq. (10) it is easy to see that for $v=0$ two boundaries exist with $\varphi_0=0$ and $\varphi_0=\pi$ (we will not discuss solutions with $\varphi_0\neq 0,\pi$, which exist for the less interesting case $h\leqslant 2\rho/\pi$ and correspond to unstable domain walls of Néel type). When $v=0$, domain walls with $\varphi_0=0$ or $\varphi_0=\pi$ correspond to rotations of the magnetization $\mathbf{M}$ in the zy plane, which includes the anisotropy axis and the direction of $\mathbf{H}_\perp$. However, at the center of the domain wall it is more energetically favorable to have $\mathbf{M}$ parallel to $\mathbf{H}_\perp$ (i.e., $\varphi_0=0$) than to have $\mathbf{M}$ antiparallel to $\mathbf{H}_\perp$ ($\varphi_0=\pi$). With increasing velocity the value of the angle $\varphi_0=\varphi_0(v)$ increases in the energetically favored wall, but decreases in the unfavored one. At a value $v=v_c$ determined by the equations

$$v_c=g M_0\sqrt{\alpha\beta}\left[\rho\ \cos\ \varphi_c+\frac{\pi h}{2}\right]\sin\ \varphi_c,$$

$$4\rho\ \cos\ \varphi_c=-\frac{\pi h}{2}\pm\sqrt{\left(\frac{\pi h}{2}\right)^2+8\rho^2}, \tag{11}$$

we have $\varphi=\varphi_c$ for both walls. The quantity v_c plays the role of a limiting wall velocity for ρ, $h\ll 1$. Analysis of the general expression Eq. (11) is quite tedious, so that we only discuss limiting behavior with respect to the parameter h/ρ here. For $h/\rho\ll 1$ the value of v_c corresponds to the Walker limit, while for $h/\rho\gg 1$ the quantity v_c increases linearly with increasing field $H_\perp$ (see Fig. 1), and

$$v_c=g M_0\sqrt{\alpha\beta}\ \frac{\pi H_\perp}{2H_a},\quad h\gg\rho,\quad H_\perp\ll H_a. \tag{12}$$

In this case the limiting velocity of a domain wall v_c exceeds the Walker limit by a factor of $\pi H_\perp/\rho H_a$. The linear dependence (12) gives a good description of experimental values of the limiting domain-wall velocity.[6] As we will see below, Eq. (12) is valid over a wide range of fields up to $H_\perp\leqslant 0.6H_a$, where the approximation $H_\perp\ll H_a$ used to derive it necessarily becomes incorrect. Note that even for $H_\perp\ll H_a$, the value of v_c in Eq. (11) exceeds the bifurcation value $v_{(-)}$, which signals changes in the type of singular points of the dynamical system (5) and (6) in the four-dimensional phase space $(\theta,\varphi,\theta',\varphi')$ corresponding to the ground state of the ferromagnet: $\sin\theta=h$, $\varphi=0$, $\theta'=0$, $\varphi'=0$ (see Ref. 14). In the Walker problem these quantities coincide. Thus, our case (of a weak transverse field) constitutes yet another example where the use of criteria based on bifurcations of singular points to find the limiting velocity of a domain wall can lead to incorrect answers for nonintegrable systems.

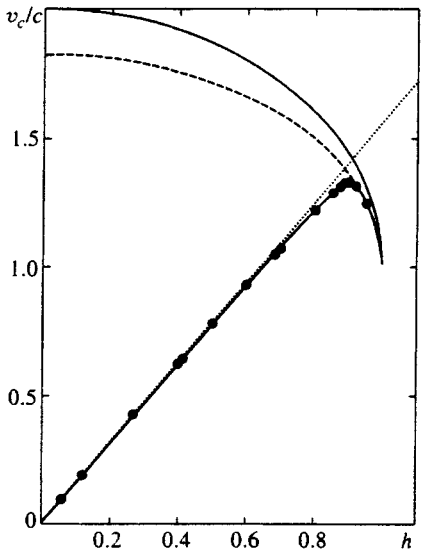

FIG. 1. Limiting velocity of a domain wall as a function of the transverse field $h=H_\perp/H_a$. Points—numerical analysis data, see below. In this figure the dotted curves denote a linear function corresponding to the low-field asymptotic behavior of Eq. (12), the dashed curves are square-root dependences corresponding to the high-field asymptotic behavior, see Eq. (24) below. The upper solid curve is the minimum phase velocity of a spin wave $v_{(+)}$.

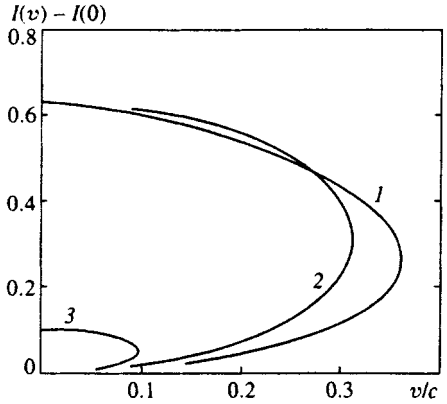

FIG. 2. The function $I(v)$ for various values of the parameters: curve *1*—$h=0.2$, $\rho=0.2$; *2*—$h=0.2$, $\rho=0$; *3*—$h=0$, $\rho=0.2$.

The approximation $\varphi(\xi)=\varphi_0=\text{const}$ is inadequate if we want to calculate the dependence of the domain wall energy on its velocity, because the integral $I(v)$ can be computed only if we take into account the fact that $\varphi(\xi)$ actually decreases at large distances from the domain wall center. However, the linearized problem with respect to ρ, h can be solved. Up to quadratic terms in the small parameters ρ and h, we can omit the $(\varphi')^2$ term and use Eq. (8) to write the integral in the form

$$I(v)=2\Delta\int_{\theta_0}^{\pi/2}\theta'\,d\theta,$$

where θ_0 is the equilibrium value of the angle θ and $\theta_0\sim h$. We will calculate this integral over two regions: $\theta\gg\theta_0$ and $\theta-\theta_0\leqslant\theta_0$.

The contribution of the second region, as is easy to see, is proportional to θ_0^2, i.e., h^2, and can be omitted. However, in the first region it follows from Eq. (8) that to accuracy h^2,

$$\Delta(\theta')=(\sin\,\theta-h\,\cos\,\varphi_0)\left(1+\frac{\rho}{2}\sin^2\,\varphi_0\right).$$

Using this expression, the integral $I(v)$ can be calculated and is elementary. In the linear approximation in ρ and h, we obtain

$$I(v)=1+\frac{1}{2}\,(\rho\,\sin^2\,\varphi_0-\pi h\,\cos\,\varphi_0), \tag{13}$$

where the value of φ_0 is determined by the velocity of the domain wall via Eq. (10). The functions $I(v)$ have similar forms for various ratios of the parameters ρ and h; see Fig. 2. They contain lower and upper branches corresponding to domain walls with different values of the angle φ, namely $0<\varphi<\varphi_0$ and $\varphi_0<\varphi<\pi$ respectively.

The approximation given above is equivalent to using the Slonczewski equations[2] for steady-state motion of a domain wall with a velocity that is not small. Since these equations are valid only in the limit h, $\rho\to0$, it is unclear why our results agree so well with the experimental values of the limiting velocity up to values of the field $h\sim0.6$, where the approximation $h\ll1$ necessarily is not satisfied. Note also that simple expressions for the domain wall energy like Eq. (13) disagree strongly with the exact results even when $h\sim0.3$. In order to unravel this problem we numerically integrated the corresponding Eqs. (5) and (6) for various values of the field h and domain wall velocity. Numerical integration showed that the condition $\varphi=\text{const}$ is rather well satis-

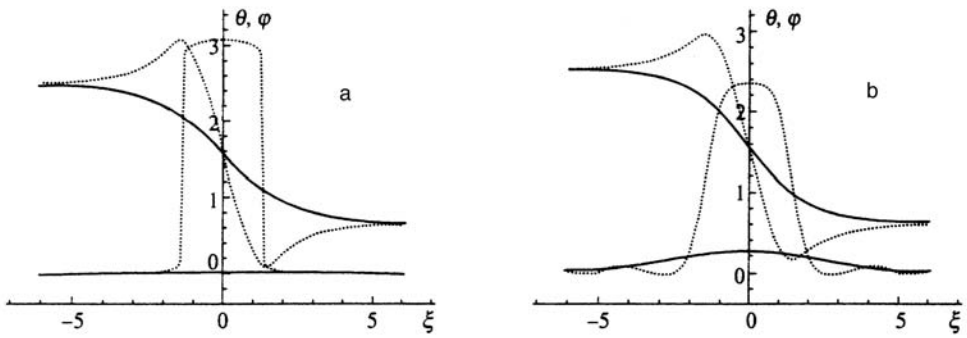

FIG. 3. Shape of the functions $\theta(\xi)$ and $\varphi(\xi)$ within a domain wall obtained by numerical integration for $h=0.6$ and various values of the domain wall velocity: a—$v=0.05c$, b—$v=0.5c$. The solid and dashed curves correspond to domain walls with small and large energies.

fied near the center of the domain wall even for values of the field and velocity that are not very small; see Fig. 3. Therefore, in order to obtain approximate expressions that are valid for moderate values of the field (h of order 0.5–0.6), we investigated the form of solutions that can be obtained using the approximation $\varphi=\text{const}$ in more detail, without assuming that the field h is small.

In experiments the field $H_\perp$ is usually large compared to the anisotropy field in the basal plane. From Fig. 2 it is clear that if the field $H_\perp$ is of the same order as the anisotropy field in the basal plane ρH_a, then the contribution of h is more important than the contribution of ρ. This is a general property for $h \geqslant \rho$; hence, in what follows we will investigate only the case $\rho=0$. The distribution $\theta(\xi)$, which can be obtained from Eq. (8), takes the following form for $\rho=0$:

$$\Delta^2[(\theta')^2+(\varphi')^2\sin^2\theta]-\sin\theta-h\cos\varphi)^2=0.$$

Integrating this expression for $\varphi=\text{const}$ yields two possible solutions. One is described by the expression

$$\tan\frac{u}{2}=\pm\sqrt{\frac{1-h\cos\varphi_0}{1+h\cos\varphi_0}}\tanh\left(\frac{\xi\sqrt{1-h^2\cos^2\varphi_0}}{2}\right),$$

$$u=\frac{\pi}{2}-\theta, \tag{14}$$

while the second is obtained from Eq. (14) by replacing tanh with coth.

When $\varphi_0=0,\ \pi$ these functions describe the exact distribution $\theta(\xi)$ for a motionless wall. If $\varphi_0=0$ we need to choose the solution (14); then as $\xi\to\pm\infty$, $\theta\to\theta_0$ and $\theta\to\pi-\theta_0$. In order to describe the less favorable wall with $\varphi_0=\pi$ we need to choose the second solution; then as $\xi\to\pm\infty$, $\theta\to-\theta_0$ and $\theta\to\pi+\theta_0$, using the notation $\sin\theta_0=h$. Although the function $\theta(\xi)$ approaches different limits as $\xi\to\pm\infty$ for $\varphi_0=0$ and $\varphi_0=\pi$, far from either type of domain wall the values of the projection of the magnetization are the same and correspond to the two possible equilibrium values of the magnetization in the ferromagnet: $m_z=\cos\theta=\pm\sqrt{1-h^2}$, $m_x=\sin\theta\cos\varphi_0=h$. Note that the solution for $\varphi_0=\pi$ can be written in a different but equivalent form by using the function $\theta(\xi)$, defined for $0<\theta\leqslant\pi/2$ and associated with $\varphi_0=\pi$, to treat the case $\theta<0$ by replacing φ_0 by 0 and $\theta(\xi)$ by $-\theta(\xi)$ (in the numerical integration we usually find exactly this type of solution; see Fig. 3a). Therefore, when $v\neq0$, i.e., for arbitrary values of φ_0, Eq. (14) gives a good approximation to the distribution $\theta(\xi)$ near the center of the domain wall.

The function $I(v)$ can be obtained in explicit form from the single assumption $\varphi=\varphi_0=\text{const}$. This function has the form

$$I(v)=\sqrt{1-h^2\cos^2\varphi_0}+h\times\left[\frac{\pi}{2}-\sin^{-1}(h\cos\varphi_0)\right]\cos\varphi_0, \tag{15}$$

where φ_0 is determined by Eq. (11). For $\rho=0$, $h\to0$ this expression becomes the asymptotic formula (13) we obtained above, while for $\varphi=0,\ \pi$ it gives exact values for the energy of a motionless domain wall. For small h Eq. (15) gives a fairly good description of the exact function $I(v)$ obtained numerically. When $h=0.6$ the discrepancy at high enough velocities does not exceed 10%, and the value of the limiting velocity obtained is accurate to about a percent; see Fig. 4.

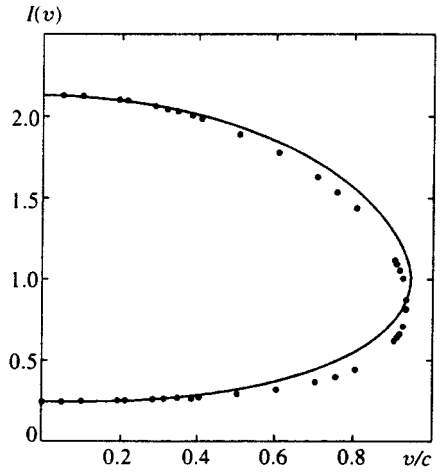

FIG. 4. The function $I(v)$ for $h=0.6$. The solid curve was plotted based on Eq. (15), the dots are the result of numerical calculations. The quantity v_c/c equals 0.9302; using Eq. (12) gives a value 0.9425.

Thus, we may conclude that although Eqs. (15) and (14) are not rigorously justified for a moving domain wall, they give a good description both of the structure and the energy of the wall (or its coefficient of friction; see below), primarily because the value of the angle φ in the solution changes slowly near the wall center. This explains why the low-field expressions for the dependence of the limiting velocity on field describe the experiments and numerical data so well up to fields $H_\perp\sim0.6H_a$.

4. BIFURCATION OF SOLUTIONS AT LARGE FIELDS

We begin our analysis of the strong-field case with the limit $H_a-H_\perp\ll H_a$. In this limit, which is reasonable when $\rho=0$, the angle $u=\pi/2-\theta$ between the ground-state magnetization of the ferromagnet and the transverse field is small, $u=\pi/2-\theta\leqslant u_0$, $u_0=\sqrt{1-h^2}\ll1$. As in the analysis of Ref. 7, we can assume that the quantities u and φ are small within the domain wall; as we will verify below, $\varphi\leqslant uu_0\ll u$. The smallness of these parameters implies that we can use the linearized versions of Eq. (6), while in Eq. (5) we keep only the simplest nonlinear term (cubic in u). Then this system of equations takes the form

$$\Delta^2\varphi''-h\varphi=vu'/c, \tag{16}$$

$$\Delta^2u''+u(u_0^2-u^2)=-v\varphi'/c. \tag{17}$$

It is easy to write the formal solution of Eq. (16) in the form

$$\varphi=-\frac{v}{c}(\mathbf{L})^{-1}u',\quad \mathbf{L}=h-\Delta^2\frac{d^2}{d\xi^2}. \tag{18}$$

Applying the operator $\mathbf{L}$ to both sides of Eq. (17) and taking Eq. (18) into account, we obtain a fourth-order differential equation for u:

$$\Delta^4 \frac{d^4u}{d\xi^4} - \Delta^2\left(1-\frac{v^2}{c^2}\right)\frac{d^2u}{d\xi^2} + \frac{1}{2}u(u^2-u_0^2) = \frac{3}{2}\left[u^2\frac{d^2u}{d\xi^2} + 2u\left(\frac{du}{d\xi}\right)^2\right], \quad (19)$$

where only the leading terms in u and u_0 are retained in each term.

We now analyze this equation. Above all, note that the width of a domain wall is large compared to the quantity Δ, which characterizes small fields. Indeed, by virtue of Eq. (19), the domain wall width is in order of magnitude equal to $(1-v^2/c^2)^{-1/2}u_0\Delta$ or $u_0^{-1/2}\Delta$ for $1-v^2/c^2 \gg u_0$ or $|1-v^2/c^2| < u_0$ respectively, and the ratio of the width of the domain wall as $h\to 1$ to its $h\approx 0$ width always contains a negative power of the small parameter $u_0=\sqrt{2(1-h)}$. The estimates we will obtain for the scale of the gradients in the domain wall show that terms on the right side of Eq. (16) are small for all values of the velocity. Neglecting these terms and introducing the variable $\psi=u/u_0$, which depends on the argument $\zeta=u_0^{1/4}\xi/\Delta$, we can write this equation in a universal form that depends only on the single parameter A:

$$2\frac{d^4\psi}{d\zeta^4} + A\frac{d^2\psi}{d\zeta^2} + \psi(\psi^2-1)=0, \quad A=\frac{v^2-c^2}{2c^2u_0}. \quad (20)$$

In this approximation, it is easy to obtain for $I(v)$

$$I(v)=2u_0^{5/2}J(A), \quad J(A)=\int_0^\infty (d\psi/d\zeta)^2 d\zeta. \quad (21)$$

Over a wide range of velocities from $v=0$ to $v\sim c$, but with $1-v^2/c^2 \gg u_0$, it is not important to include the fourth derivative in Eq. (20). This leads us to the Lorentz-invariant equation of the φ^4 model,[7] which has the following form in the case of interest to us, i.e., a simple wave:

$$\Delta^2\left(1-\frac{v^2}{c^2}\right)\frac{d^2u}{d\xi^2} - \frac{1}{2}u(u^2-u_0^2)=0. \quad (22)$$

Using the well-known kink solution to this equation, it is easy to write the distribution of magnetization in a moving domain wall as

$$\theta=\frac{\pi}{2}-\sqrt{1-h^2}\,\tanh\left[\frac{y-vt}{\Delta(v)}\right],$$
$$\varphi=\frac{v(1-h^2)(c^2-v^2)^{-1}}{\cosh^2[(y-vt)/\Delta(v)]}, \quad (23)$$

where $\Delta(v)=\Delta\sqrt{1-h^2}/\sqrt{1-v^2/c^2}$ is the thickness of the domain wall in this approximation, i.e., for $(1-v^2/c^2)\gg\sqrt{1-h^2}$. It is easy to see that the condition $\varphi \ll \pi/2-\theta \ll 1$ is satisfied for $(1-v^2/c^2)\gg\sqrt{1-h^2}$. From Eq. (23) it follows that as the domain wall velocity increases its thickness undergoes a Lorentz contraction and decreases to a value of order $[\alpha^2/\beta^2(1-h)]^{1/4}$ when $(1-v^2/c^2)\sim\sqrt{1-h^2}$. The Lorentz invariance of the equation of the φ^4 model Eq. (22) implies that the dependence of the domain wall energy on its velocity is given by the expression

$$I(v)=(\sqrt{1-h^2}/3)/\sqrt{1-v^2/c^2} \quad \text{for } h\approx 1,$$
$$(1-v^2/c^2)\gg\sqrt{1-h^2}.$$

As $v\to c$ these simple expressions become invalid, and we are forced to analyze the fourth-order Eq. (20) in full. Unfortunately, this can only be done numerically. However, the problem is simplified by the fact that the limiting velocity as $h\to 1$ is close to c, differing from it by a quantity of order $c\sqrt{1-h^2}$. We can expect that the value of v_c lies in the interval between the bifurcation values $v_{(+)}$ and $v_{(-)}$, where $v_{(\pm)}=c(1\pm\sqrt{1-h^2})$. In any case, it cannot exceed the phase velocity of spin waves $v_{(+)}$. Therefore, it is sufficient to investigate the fourth-order Eq. (20) numerically for values of the parameter A in the interval from $A=1$ to $A=-1$, corresponding to $v=v_{(+)}$ and $v=v_{(-)}$ respectively. Since Eq. (20) depends on v and h only through the universal parameter A, data obtained for one value of A determine the structure of the domain wall for various values of v and field $H_\perp$.

Numerical integration of Eq. (20) shows that for $A<0$ (i.e., $v\leqslant c$), only solutions of the ''ordinary'' domain wall type exist; see, e.g., Eq. (23), for which the angle θ varies monotonically. When $A>0$ $(v>c)$, the asymptotic form of this type of solution becomes nonmonotonic. This nonmonotonicity arises from a change in the nature of the singular points that describe the ground state of the ferromagnet. In principle, the $A>-1$, i.e., $v>v_{(-)}$, solutions should also be nonmonotonic, but here the effect is not so obvious, and is not apparent when the range of numerical integration is finite. For $A>0$ another type of solution also appears with higher energy. For values of A that are not too large, this type can be treated as a cluster of five domain walls, see Fig. 5a. Note that here, in contrast to the small-field case of a higher-energy domain wall, there is a nonmonotonicity in the coordinate dependence of the magnetization components. As A increases, the shapes of the various walls approach one another (Fig. 5b), until a critical value $A=A_c$ is reached at which the solutions coincide (Fig. 5c; A_c lies between 0.8229 and 0.8230). This scenario for reaching the limiting velocity is characteristic only of nonintegrable dynamic systems, as are solutions consisting of domain wall clusters type or domain walls with oscillatory behavior, and are discussed in Refs. 9, 15, and 16.

Thus, the limiting velocity of a domain wall can be found from the condition $A=A_c$. It is determined by the expression

$$v_c=c(1+A_cu_0)\simeq c(1+0.823\sqrt{1-h^2}), \quad h\to 1, \quad (24)$$

and decreases with increasing field; see Fig. 1 above.

In order to find the regions of applicability of these asymptotic dependences, we compare them to the results of a numerical analysis of the Landau–Lifshitz equations for large values of the field h. Here convergence to the asymptotic limit is slower than in the small-field case. In fact, the function $v_c(h)$ does not even decrease until $h>0.89$, and the asymptotic form (24) is approached only at $h>0.95$; see Fig. 1 above.

The asymptotic dependences of the limiting domain wall velocity on the transverse field $H_\perp$ obtained here reveal several types of behavior with increasing field: linear increase for small fields, as in Eq. (12), and square-root behavior for large fields, as in Eq. (24). As we have noted, the domain

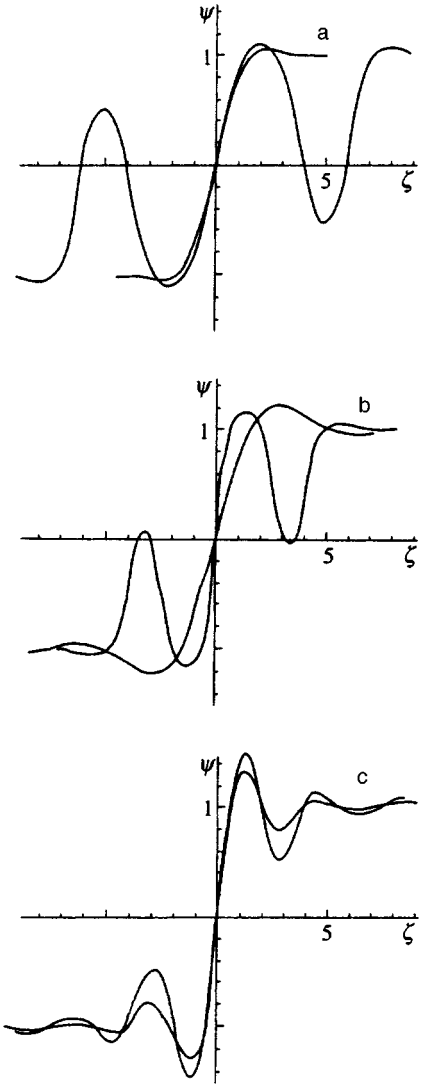

FIG. 5. Form of solution for various values of the parameter A: a—$A=0$ ($v=c$); b—$A=0.5$; c—$A=0.8$.

wall velocity is limited by the merging of the upper and lower branches at the bifurcation value $v=v_c$. However, the domain walls for the upper and lower branches are qualitatively different for small and large fields. As we discovered in the previous sections, the low-field and high-field asymptotic expressions are correct for $h<0.6$ and $h>0.95$ respectively. There remains the question of how the domain wall behaves in intermediate fields $0.6<h<0.9$, and how the transition from one type of behavior to another takes place as the field increases. Although the problem exhibits no universality in the general case, forcing us to investigate solutions that depend on two parameters, i.e., v and h, we will show that a fairly complete picture can be given without a particularly high computational cost.

As we noted above, two types of wall exist for small fields, for which the function $I(v)$ is given by a closed curve containing two branches—an upper and a lower; see Figs. 2 and 4 above. In fact, nonintegrable systems exhibit a sequence of such curves that lie on top of one another.[16] The top curves correspond to solutions in the form of domain-wall clusters, with energies that are several times the energy of a domain wall from the lower curves; therefore, they usually are not discussed.

However, in our case the situation is otherwise. Larger magnetic fields make walls with antiparallel orientation of $\mathbf{H}_\perp$ and $\mathbf{M}(0)$, which are characteristic of weak fields, even less favorable energetically. However, such fields also decrease the amplitude of the oscillations in the magnetization of the domain wall until such domain walls become advantageous. Hence, increasing the field causes the upper branch of the ''old'' curve to rise, while the upper ''new'' curve drops. At sufficiently large fields it is necessary to study all of these solutions. In particular, as Fig. 6a shows, for values $h\geqslant 0.8$, plots of the function $I(v)$ for the new and old solutions intersect near v_c. At points of intersection, the walls that correspond to the two different curves differ strongly from one another, i.e., walls of different types cannot link with one another. It is typical of the old solutions that the nonmonotonicity of the function $\theta(\xi)$ is weak, and the value of φ at the center of a domain wall corresponding to the upper branch is fairly large. The new solutions are characterized by larger nonmonotonicity of the function $\theta(\xi)$.

As the field increases, the two upper branches of the function $I(v)$ approach one another. When the characteristic value $h=h^*\approx 0.88$ is reached, they merge (see Fig. 6b), and then the curves break up and the branches reconnect such that the upper and lower branches for both functions ''stick together.'' At still larger fields, the curves diverge, and a newly formed upper curve, in which the value of φ at the center of the wall for both domain walls is not small, rapidly moves upward; see Fig. 6c.

We illustrate the system behavior near $h=0.88$; see Fig. 7. The two low-energy domain walls belonging to different branches for $h<h^*$ are qualitatively similar to the domain walls obtained in the high-field limit (as shown in Figs. 7a and 5 above). In these walls, the angle φ is relatively small. The angle θ in the domain wall with smaller energy varies essentially monotonically with the coordinate ξ, while $\theta(\xi)$ for the higher-energy wall (of this pair) exhibits strong oscillations, but basically between the values $\pi/2$ and θ_0, i.e., there is not much movement ''downward.'' For the pair of domain walls with larger energy the situation is opposite: they are more similar to domain walls at low fields (see Fig. 7b and Fig. 3 above). In these two domain walls the angle φ is large, and in the center of both domain walls it is close to $\pi/2$ and the amplitude of the change in φ considerably exceeds the amplitude of the change in θ. In these walls we also observe an abrupt dip in $\theta(\xi)$ when $\xi\leqslant\Delta$; in the neighborhod of this dip, φ goes to zero.

Thus, the four domain walls observed in the transition region at $h\sim h^*$ are qualitatively similar to the ''low-field'' and ''high-field'' domain walls described above. The wall with almost monotonic variation of θ and φ always has the lesser energy. The cases $h>h^*$ and $h<h^*$ differ in that for relatively small fields a domain wall with monotonic variation of θ and φ belongs to the same curve as the low-field wall and continuously transforms into it, while for larger fields it continuously transforms into a high-field domain wall.

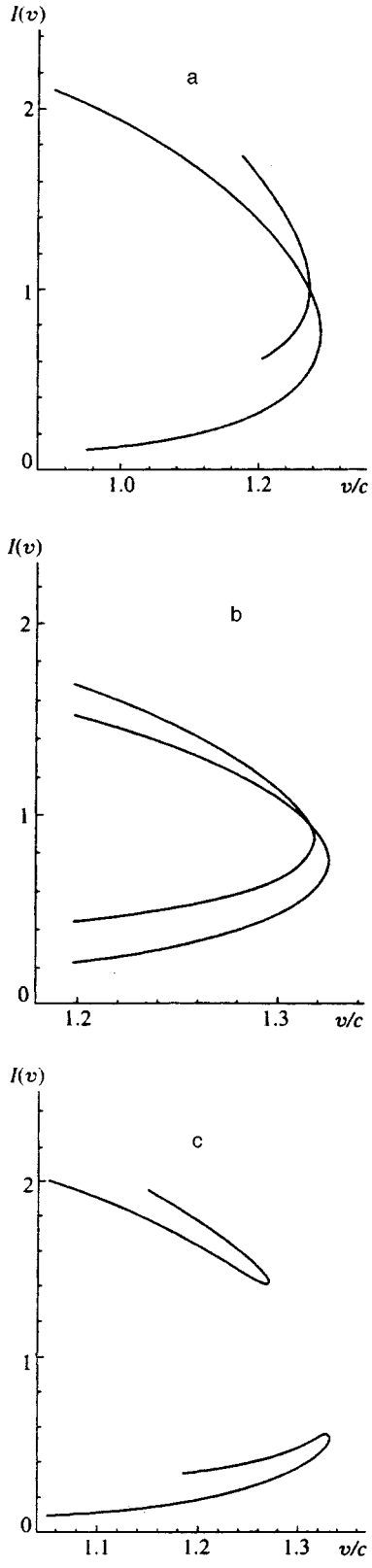

FIG. 6. The function $I(v)$ for various values of h: a—$h=0.85$; b—$h=0.88$; c—$h=0.9$.

Thus, our analysis of the Landau–Lifshitz equation not only demonstrates the validity of the approximate solutions we have obtained for large and small fields, but also allows us to establish their range of applicability. In the general case $h\sim h^*$ and $v\leqslant v_c$, four types of domain walls exist; the approximate solutions describe only two of them—those that possess the smaller energies. The two walls with larger energies do not appear within our approximate treatment for the reasons discussed above. At small fields these walls are characterized by nonmonotonic changes in $\theta(\xi)$ and $\varphi(\xi)$, and necessarily cannot be described by trial functions of type $\varphi(\xi)\approx\varphi_0=\text{const}$. At large fields the higher-energy walls are not low-amplitude walls.

The presence of several types of domain walls makes the question of their stability a crucial one. It is likely that only those walls with minimum energy are stable, as in the Walker model; see Refs. 1 and 4. More detailed analysis of this problem is beyond the scope of this paper.

5. BRAKING AND THE CHARACTER OF FORCED MOTION OF DOMAIN WALLS

Calculation of the braking force on a domain wall addresses an important problem in the physics of magnetism, namely that of relaxation of elementary excitations in magnets, both linear (e.g., magnons) and nonlinear (e.g., topological solitons that describe the domain walls). Usually, the braking of domain walls is studied using the Landau–Lifshitz equations with a relaxation term of Hilbert form. However, this approach is not always correct. In particular, it is known that its use in calculating the decay rate for attenuation of magnons $\gamma(k)$ leads to qualitative contradictions with experiment and the results of microscopic theory.

In describing the dynamics of domain walls, an important parameter is the force of viscous friction f that acts on a domain wall moving with velocity v per unit wall area, i.e., $f=-\eta(v)v$, where $\eta(v)$ has the sense of a nonlinear coefficient of viscosity. Calculation of $\eta(v)$ for domain walls based on the Landau–Lifshitz equation with the standard relaxation term leads to a number of inconsistencies with experiment (see the discussion of this question in Refs. 2, 8, 17–19, and 21).

In Refs. 17 and 18, Bar'yakhtar *et al.* proposed a generalized phenomenological theory of relaxation in ferromagnets based on taking the dynamical symmetry of the magnet into account. In this theory, relaxation terms of various kinds are introduced (exchange and relativistic) that lead to correct expressions for the dependence of the magnon decay rate $\gamma(k)$ on wave vector k, and make it possible to describe a number of experiments on the dynamics of magnetic solitons. The authors of Ref. 12 noted that their measurements of the dependence of domain wall motion on the magnetic field $H_\perp$ perpendicular to the axis of easy magnetization could be quantitatively described using this theory (a separate calculation of the viscosity coefficient for small velocities was given in Ref. 10).

The generalized theory is based on the Landau–Lifshitz equation (1) for the magnetization with a relaxation term of the form

$$\mathbf{R}=gM_0\Lambda\mathbf{F}-\lambda_e a^2 gM_0\nabla^2\mathbf{F}.$$

Here λ_e is the exchange relaxation constant and Λ is a tensor of relativistic relaxation constants; see Refs. 17 and 18 and below. For a rhombic ferromagnet the tensor Λ can be cho-

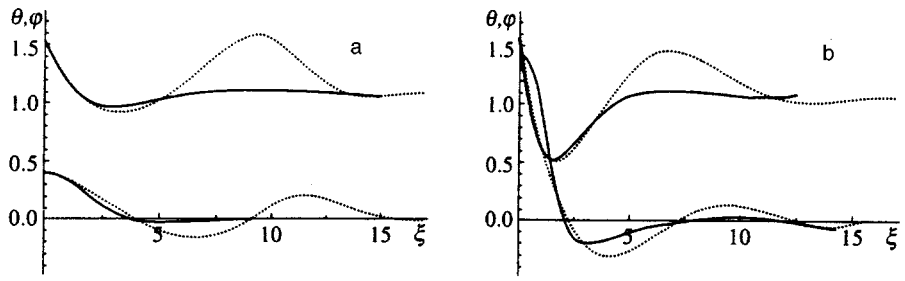

FIG. 7. The functions $\theta(\xi)$ and $\varphi(\xi)$ within a domain wall obtained by numerical integration for $h=0.88$ and $v=1.2c$. Only half of each curve is shown; the second half is easily reproduced by following the rules $\theta(-\xi)=\pi-\theta(\xi)$, $\varphi(-\xi)=\varphi(\xi)$ (see Figs. 3 and 5 above). Plot a corresponds to a pair of domain walls with lower energy, plot b to a pair with higher energy. In each figure the solid and dotted curves correspond to domain walls from a given pair with low and high energies respectively.

sen to be diagonal, $\Lambda=\mathrm{diag}(\lambda_x,\lambda_y,\lambda_z)$. For a uniaxial ferromagnet $\lambda_x=\lambda_y=\lambda$. When only intrinsic relaxation processes are included, the z-projection of the total magnetization is conserved, and hence $\lambda_z=0$.[18] If, however, impurity-induced relaxation is included, the tensor Λ for ferrite garnets has cubic symmetry, i.e., $\lambda_x=\lambda_y=\lambda_z$.[20] Although the magnetic symmetry is reduced in large magnetic fields, in the interest of simplicity we choose to limit ourselves to tensors Λ that are diagonal,

$$\Lambda=\lambda\ \mathrm{diag}(1,1,\varepsilon), \tag{25}$$

while not yet specifying the value of the constant ε.

The rate of energy dissipation of a domain wall, or the coefficient of nonlinear viscosity, is directly determined by a dissipation function

$$dE/dt=-2Q, \quad \eta(v)=2Q/v^2.$$

Taking Eq. (1) into account, we can write the dissipative function of the ferromagnet as a functional of the effective field $\mathbf{F}$[17,18]:

$$Q=\frac{1}{2}\,gM\int d\mathbf{r}\{\lambda_{ik}F_iF_k+\lambda_e a^2(\nabla\cdot\mathbf{F})^2\}. \tag{26}$$

In order to analyze the braking of domain walls we must derive the dissipative function Q in terms of the time derivative of the magnetization. Therefore, in order to compute $\eta(v)$ we must express $\mathbf{F}$ in terms of $\mathbf{m}$ and its derivative. By virtue of Eq. (1), the effective field $\mathbf{F}$ can be written in the form of two terms that are respectively perpendicular and parallel to the magnetization. The value of the parallel component of $\mathbf{F}$ is associated with changes in the length of the magnetization vector $\mathbf{M}$, and in principle it cannot be found from the dissipationless Landau–Lifshitz equation. However, if we assume that (*i*) the exchange interaction is sufficiently small ($\lambda_e\ll\lambda\Delta^2/a^2$), and (*ii*) the domain wall velocity is less than the characteristic velocity v^* ($v\ll v^*=g\lambda M\Delta/\chi$), where Δ is the width of the domain wall, and χ is the longitudinal susceptibility of the ferromagnet (the susceptibility for paramagnetism), then we can write the dissipation function explicitly as a quadratic form involving $\partial m_i/\partial t$ (see Ref. 21 for details).

Let us discuss how realistic these inequalities are. Approximation (*i*) is realistic for a broad class of ferromagnets, including weakly anisotropic yttrium iron garnet and epitaxial films of ferrite–garnets used in magnetic-bubble structures. For these materials, the physical reasons for the smallness of the exchange interaction are different: for yttrium iron garnet the reason is the smallness of the gradient of the magnetization in the domain walls, while for magnetic bubble materials it is that the constant λ is large; see Refs. 11 and 19.

As for inequality (*ii*), the quantity v_c can be rewritten in the form $v_c=c(\lambda/\beta\chi)[\Delta(h)/\Delta(0)]$. For small h this quantity contains the ratio of the small parameters λ and $\chi\beta$; thus, we may assume that it is of order c, while $v_c\ll c$. For large $h\sim 1$, $v_c=c$, but the right-hand terms of both inequalities contain powers of the large parameter $\Delta(h)/\Delta$. Therefore, we may assume that neither inequality imposes strong restrictions on the applicability of the theory we have developed to real ferromagnets.

When these two inequalities hold, the dissipation function can be written as a sum of three terms:

$$Q=Q_r+Q_e+Q_\chi, \tag{27}$$

each of which admits a simple physical interpretation. In what follows, we will use the fact that $\eta_\alpha=Q_\alpha/v^2$ (where $\alpha=r$, e, or χ), while in the domain wall $\partial(\theta,\varphi)/\partial t=-v(\theta,\varphi)'$ (the dashes denote derivatives with respect to ξ), to derive expressions for the contributions of these three terms to the viscosity coefficients η_r,η_e,η_χ.

The quantity η_r can be interpreted as the direct contribution from relativistic relaxation. In angular variables for the magnetization vector, and when we choose the tensor Λ in the form $\lambda\,\mathrm{diag}(1,1,\varepsilon)$, the quantity η_r can be reduced to the form

$$\eta_r=\frac{\lambda M_0}{g}\int d\xi\left\{(\theta')^2+\sin^2\theta\,\frac{\varepsilon}{\sin^2\theta+\varepsilon\cos^2\theta}\,(\varphi')^2\right\}. \tag{28}$$

For $\varepsilon=1$ and $\lambda_{ik}=\lambda\delta_{ik}$, the expression $(\theta')^2+\sin^2\theta(\varphi')^2$ is under the integral sign, and this expression reduces to a dissipative function of Hilbert form.

The exchange contribution is defined by the expression $\eta_e=(M_0\lambda_e a^2/g)\int d\xi(\mathbf{m}''\cdot\mathbf{m})$. Its form in angular variables, which is rather cumbersome, has appeared in many papers, and we will not discuss it here; see Refs. 18 and 21. Finally, the third term η_χ determines the contribution of changes in the length of the magnetization vector to the dissipation of a soliton.[19,21] The explicit expression for η_χ in our model can be written in the form

$$\eta_\chi=\frac{M_0}{g}\frac{\beta^2\chi^2}{\lambda}\int d\xi\left\{\frac{1}{\sin^2\theta+\varepsilon\cos^2\theta}\right\}\times\{[2\sin^2\theta(1+\varepsilon\sin^2\varphi)-3h\sin\theta\cos\varphi]'\}^2. \tag{29}$$

This contribution is nonzero only for finite values of the longitudinal susceptibility χ of the ferromagnet. The role of ''effective relaxation constant'' in this term is played by the quantity $\beta^2\chi^2/\lambda$.

Using these general expressions, we can compute the dependence of the partial viscosity coefficients on the domain wall velocity. The specific calculations can be carried out analytically in the limiting cases of small and large fields, i.e., h, $\rho\ll 1$ and $h\sim 1$. For an exact analysis we compute $\eta_\alpha(v)$ using the data from numerical integration of the complete system of Landau–Lifshitz equations. Once the function $\eta(v)$ is found, the dependence of the velocity of forced motion of the domain wall on the forcing field H_z is easily obtained by setting the frictional force equal to the magnetic pressure force: $v\,\eta(v)=H_z M_0\cos\theta_0$. Solving this equation, it is easy to construct the function $v=v(H_z)$ for various values of the problem parameters.

For small fields $H<0.5H_a$, the integrals can be calculated by the same approach as we used above to calculate the integral $I(v)$; see Eqs. (13)–(15). Since the contribution of h is more important than the contribution of ρ for $h\sim\rho$, we give the expressions for $\rho=0$ here. After some simple transformations the contributions of all three mechanisms can be written in the form

$$\eta_r(v)=\frac{2\lambda M_0}{g\Delta}I_r(v),\quad \eta_e(v)=\frac{2\lambda_e a^2 M_0}{g\Delta^3}I_e(v),$$

$$\eta_\chi=\frac{2M_0\beta^2\chi^2}{g\lambda\Delta}I_\chi(v), \tag{30}$$

where $I_\alpha(v)$ are dimensionless functions similar to Eq. (7). In our chosen approximation $\varphi=\text{const}$, the coefficients η_r and η_e do not depend on the parameter ε in the tensor Λ. The quantity $I(v)$ coincides with what we calculated above; see Eqs. (13) and (15). For $\varphi=\text{const}$ the general expression for the contribution of exchange relaxation leads to the form

$$I_e(v)=\Delta^3\int_0^\infty(\theta'')^2 d\xi,$$

which yields

$$I_e(v)=\frac{1}{3}(1-h^2)^{3/2}-h\cos\varphi_0\left[\frac{\pi}{2}-\theta_0-h(1-h^2)^{1/2}\right],$$

where φ_0 is determined by the velocity of the domain wall (see Eq. (10) and Eq. (10′) below), and $\sin\theta_0=h$.

The form of $I_\eta(v)$ depends heavily on the ratio of λ and λ_z. In the limits $\lambda_z\ll\lambda$ and $\lambda=\lambda_z$, which we expect to hold for small fields when intrinsic or impurity relaxation processes dominate, respectively, we have

$$I_\chi(v,\varepsilon=0)=\frac{16}{3}c_0^3-20\cos\varphi_0(hu_0-c_0)$$

$$-33h^2\cos^2\varphi_0\left[c_0+\ln\tan\left(\frac{\theta_0}{2}\right)\right]$$

$$-9h^2\cos^3\varphi_0(c_0-hu_0),$$

$$I_\chi(v,\varepsilon=1)=\frac{16}{15}c_0^3(5-3c_0^2)-20h\cos\varphi_0\times[u_0+hc_0(2c_0^2-1)]+11c_0^3h^2\cos^2\varphi_0$$

$$-\frac{9}{2}h^3\cos^2\varphi_0(u_0-hc_0), \tag{31}$$

where $u_0=\pi/2-\theta_0$, $c_0=\cos\theta_0=(1-h^2)^{1/2}$. As $h\to 0$, it is easy to obtain

$$I_\chi(v,\varepsilon=0)=\frac{16}{3}\left(1-\frac{15}{8}\pi h\cos\varphi_0\right),$$

$$I_\chi(v,\varepsilon=1)=\frac{32}{15}\left(1-\frac{75}{64}\pi h\cos\varphi_0\right),$$

and the expressions for η_χ become those obtained previously in Ref. 8, while for $v\ll v_c$ and $\varphi_0=0$ they become the expressions for the linear mobility of a domain wall.[10]

In all of the expressions in (30) and (31), the velocity of the domain wall appears only in the quantity $\cos\varphi_0$, from which by virtue of Eq. (10) it is easy to obtain

$$\cos\varphi_0=\pm\sqrt{1-v^2/v_c^2}, \tag{10′}$$

where the lower sign corresponds to a stable domain wall with $0<\varphi_0<\pi/2$, and the upper sign to an unstable domain wall with $\pi/2<\varphi_0<\pi$. These dependences of $\eta_\alpha(v)$ on the velocity v for $h<0.5$ are in good agreement with the data from numerical calculations.

It is not difficult to verify that for $h\ll 1$ the function $v=v(H_z)$ is essentially linear up to values near v_c, and the contributions from exchange and relativistic relaxation are indistinguishable. This function has a characteristic ''lobe'' shape, whose width increases with increasing transverse field. The lobe is narrowest for the quantity $\eta_r(v)$, while for $\eta_\chi(v)$ it is broadest. As in the Walker model, motion of the domain wall is possible only when the forcing field H_z does not exceed a certain critical value H_c. To leading approximation in h, $H_c=v_c\eta(0)/2M_0$, i.e., H_c, like the limiting velocity v_c, increases linearly with increasing transverse field h.

In the limit $h\to 1$, or $u_0=\sqrt{2(1-h)}\ll 1$, we will restrict our analysis to the Lorentz-invariant region $(1-v^2/c^2)\gg u_0$. For the three contributions η_r, η_e, and η_χ, to leading approximation in the small parameter u_0 we obtain expressions that can be written in a universal form (the universality is due to the fact that all the velocity dependence results from the Lorentz contraction of the domain wall):

$$\eta_\alpha=C_\alpha(1-h^2)^{n/2}/(1-v^2/c^2)^{k/2}, \tag{32}$$

where $n=1$ for η_r, $n=3$ for η_e and η_χ; $k=1$ for η_r and η_χ; and $k=3$ for η_e. When $\varepsilon=1$, which is natural when $h\approx 1$, the constants C are given by

$$C_r=\frac{4\sqrt{2}M_0\lambda}{3g}\left(\frac{\beta}{\alpha}\right)^{1/2},\quad C_e=\frac{8\sqrt{2}M_0\lambda_e a^2}{8g}\left(\frac{\beta}{\alpha}\right)^{3/2},$$

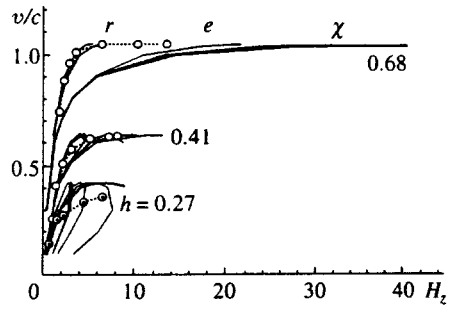

FIG. 8. Dependence of the velocity of forced motion of domain walls on the forcing field H_z (in arbitrary units) for several values of the transverse field h (shown to the right of the corresponding groups of curves). The circles are experimental data from the authors of Ref. 6.

$$C_\chi = \frac{8\sqrt{2}M_0\beta^2\chi^2}{15g\lambda}\left(\frac{\beta}{\alpha}\right)^{1/2}.$$

Note that all three contributions have different dependences on the characteristic parameters $(1-h^2)$ and $(1-v^2/c^2)$. As $H\to H_a$ and for moderate velocities, the direct term from relativistic relaxation η_r should become the dominant one (the contributions of η_χ and η_e contain one more power of the small parameter $(1-h^2)$ than η_r does). In the limit $v\to c$, the contribution due to exchange relaxation should dominate since this term increases more abruptly; see Eq. (32). When the contributions η_r or η_χ are dominant, the dependence of v on the forcing field H_z has a form that is standard for Lorentz-invariant models,

$$v(H_z) = \frac{\mu H_z}{\sqrt{1+(\mu H_z/c)^2}}, \tag{33}$$

where μ is the linear mobility of the domain walls. If, however, the contribution from exchange relaxation dominates, the function $v=v(H_z)$ saturates more rapidly.

Estimates of the discarded terms in deriving Eqs. (32) or (33) show that these equations are adequate for the problem within a fairly narrow range of $H_\perp - H_a$, specifically when $(H_\perp - H_a)/H_a \leqslant 10^{-2}$, for which $u_0 \leqslant 0.15$ and can actually be considered small. Experiments in this range of fields are quite complicated, primarily due to the low Faraday contrast of the domains. When the inequality $(1-v^2/c^2) \gg u_0$ is violated, the results (32) and (33) change. The primary change is that the functions $\eta_\alpha(v)$ have finite values as $v\to v_c$. This gives rise to a limiting value for the field $H_c = v_c\eta(v_c)/2M_0$, such that for $H_z \geqslant H_c$, stationary motion of the domain walls is impossible.

Let us discuss the corresponding behavior predicted by the complete set of Landau–Lifshitz equations. The corresponding functions $f=f(v)$ or $v=v(H_z)$ differ strongly, depending on which of the contributions dominate. The shapes of these curves are also different for different values of the transverse field $H_\perp$; see Fig. 8. For clarity these functions are plotted for that choice of effective constants λ, $\lambda_e a^2\beta/\alpha$, and $\beta^2\chi^2/\lambda$ that produces the same values of the mobility on the linear portions. The exception is the case $h=0.27$, for which we plotted the dependence only for the necessarily stable domain wall with the lowest energy.

The theoretical curves exhibit the same tendencies we observed in analyzing the limiting cases. For small fields the functions are close to linear, and saturate strongly for large fields. However, in contrast to the ''relativistic'' limit of type Eq. (33), a limiting value for the field H_c always appears for large values of the field, which is determined by the maximum value of the corresponding braking force.

The curves for the contribution η_r lie farthest to the left, followed by the curve for the contribution η_e. The contribution from η_r is essentially independent of ε. The curve arising predominantly from the contribution of η_χ is farthest to the right. Whereas for small velocities η_χ is essentially independent of ε, values of the critical field for $\varepsilon=1$ and $\varepsilon=0$ differ by almost a factor of one and a half. This figure also shows the experimental points obtained in Ref. 6, which we discuss below.

6. CONCLUDING REMARKS AND DISCUSSIONS OF EXPERIMENTS

Let us consider the available experimental data. The dependence of the domain wall velocity on the forcing field for moderate domain wall velocities (up to the limiting velocity) and values of the in-plane field up to $0.7H_a$ were investigated in Ref. 6. The theory we have constructed here is in fairly good agreement with these experimental data. In particular, over this range of fields, the limiting velocity v_c increases linearly with increasing $H_\perp$, and its value is well described by Eq. (12).

The authors of Ref. 12 noted that their experimental data on the dependence of the linear mobility of domain walls on the transverse field are well described by generalized relaxation theory up to $H_\perp < 0.5H_a$. However, in order to confirm the nonlinear theory for motion of domain walls it is important to ask a subtler question about the braking of domain walls at moderate velocities. In this case experiment yields more information, because we can analyze the shape of the curves $v=v(H_z)$, not just the value of linear mobility. In agreement with the theory developed above, the shapes of these curves differ strongly for various values of the transverse field $H_\perp$, and also for various dominant relaxation processes; see Fig. 8 above.

In order to compare theory with experiment, the theoretical curves $v=v(H_z)$ for the various viscosity contributions are plotted in Fig. 8 for the same values of the transverse field $H_\perp$ as in the experiments of Ref. 6 ($H_\perp/H_a = 0.68$, 0.41, and 0.27). Comparing these data, it is easy to see that including only relativistic relaxation, which is equivalent to using the standard relaxation term in Hilbert form, cannot describe the shape of the curve even qualitatively. In fact, whereas for small velocities the experimental points lie almost on a straight line corresponding to the contribution η_r, as the forcing field H_z increases we observe clearcut saturation. Since for $H_\perp/H_a = 0.27$ the authors of Ref. 6 did not observe saturation, we will discuss only the stronger fields in what follows.

Steady-state motion is observed up to values of the field H_z that are considerably larger than we would expect from a theory that takes only relativistic relaxation into account. On the other hand, when only the terms η_e and η_χ are included,

the shape of the curves patently fails to agree with experiment. The observed saturation of the function $v(H_z)$ can be explained by taking into account the contribution from the change in modulus of the magnetization η_χ, and assuming that the corresponding effective relaxation constant $\lambda_\chi = \beta^2\chi^2/\lambda$ is several times smaller than the relativistic constant λ. Then the theoretical curve in the weak-field region will be essentially the same as the curve for the contribution η_r, which gives a good description of experiment in this region, while the saturation of the function $v(H_z)$ finds explanation in the behavior of the contribution η_χ, which increases rapidly with increasing velocity. Based on the shape of the curve for the exchange relaxation contribution, it is clear that including it may not improve agreement between theory and experiment, because for saturation we must assume that $\lambda \sim \lambda_e a^2\beta/\alpha$, while in this case the disagreement will be appreciable at low velocities. We were unable to make a more detailed fit due to insufficient experimental data.

Thus, our generalized relaxation theory, taking into account the contributions from relativistic relaxation η_r and varying modulus of the magnetization η_χ, enables us to describe the experimental dependence of the domain wall velocity on the forcing field for various values of the transverse field over the entire range of velocities up to the limiting velocity. Analysis of these data for the sample used by the authors of Ref. 6 indicates a transition from one relaxation mechanism to another as the velocity increases. At higher velocities the contribution to the relaxation from the change in the length of the magnetization vector under the action of the effective field created by the moving domain wall turns out to be important, acting to relax $|\mathbf{M}|$ further towards its equilibrium value M_0. The importance of these two contributions (η_r and η_χ) was noted by the authors of Refs. 11, 12, and 19 (in these papers the authors obtained values other than those given here for the constants λ and $\beta^2\chi^2/\lambda$, but this is not surprising since the quantity $\beta\chi/\lambda$ can change markedly from sample to sample).

Our conclusion that the change in modulus of the magnetization makes an important contribution to the braking of a domain wall may appear strange, since domain wall dynamics have been investigated for many years without any previous perception that this mechanism is necessary. However, domain wall motion has usually been investigated in the linear regime. (Nonlinearity in ferrite–garnets without a transverse field is usually associated with ''twisting'' of the domain walls,[2] and is not discussed here). In this case the experiments were designed so that the viscous friction coefficient of domain walls (or the effective relaxation constant λ_{eff}) enters into the theory as an empirical parameter. Under these conditions, any lack of agreement between theory and experiment is apparent only in the fact that the effective constant λ_{eff} turns out to be different when measured in different ways (based on mobility of domain walls versus, e.g., measurements of the width of ferromagnetic resonance lines). Such differences are well known from studies of various ferromagnets.[2] It is a more complicated task to describe the shape of the curve $v(H_z)$, especially when the structure of the domain walls in the film changes appreciably during the experiment (for example, in a ferromagnet when the transverse field changes, or in magnets in the region of spin reorientation), which can change the ratio of the various contributions. In this case the first-principles inclusion of real relaxation mechanisms becomes crucial.

The authors thank M. V. Logunov, V. V. Randoshkin, and A. Ya. Chervonenkis for providing the experimental data, K. A. Safaryan for his help, and V. G. Bar'yakhtar and V. M. Eleonskiĭ for useful discussions. The work of one of the authors (B.I.) was partially supported by a grant from the Science and Technology Center of the Ukraine (No. 300) and the Fund for Fundamental Research of the Ukraine (No. 2.4./27).

[1] V. G. Bar'yakhtar, M. V. Chetkin, B. A. Ivanov, and S. N. Gadetskii, *Dynamics of Topological Magnetic Solitons*, Springer Tracts in Modern Physics, Vol. 120, Springer-Verlag, Berlin (1994).
[2] A. P. Malozemoff and J. C. Slonczewski, *Magnetic Domain Walls in Bubble Materials*, Academic Press, New York (1979).
[3] V. G. Bar'yakhtar and B. A. Ivanov, Sov. Sci. Rev. Sec. A-Phys., I. M. Khalatnikov ed., Vol. 16, No. 3 (1992).
[4] A. M. Kosevich, B. A. Ivanov, and A. S. Kovalev, Phys. Rep. **194**, 117 (1990).
[5] K. Blanke, B. Luhrman, V. Wallenhorst, H. Dotsch, and W. Tolksdorf, Phys. Status Solidi A **124(a)**, 359 (1991).
[6] M. V. Logunov, V. V. Randoshkin, and A. Ya. Chervonenkis, Pis'ma Zh. Tekh. Fiz. **15**(9), 64 (1989) [Sov. Tech. Phys. Lett. **15**, 731 (1989)].
[7] B. A. Ivanov, V. P. Krasnov, and E. V. Tartakovskaya, Pis'ma Zh. Tekh. Fiz. **13**, 341 (1987) [Sov. Tech. Phys. Lett. **13**, 139 (1987)].
[8] B. A. Ivanov and K. A. Safaryan, Fiz. Nizk. Temp. **18**, 722 (1992) [Low Temp. Phys. **18**, 511 (1992)].
[9] B. A. Ivanov, N. E. Kulagin, and K. A. Safaryan, Physica B **202**, 193 (1994).
[10] B. A. Ivanov and K. A. Safaryan, Fiz. Tverd. Tela (Leningrad) **32**, 2034 (1990) [Sov. Phys. Solid State **32**, 1183 (1990)].
[11] B. A. Ivanov, A. Stankiewic, A. Maziewski, N. L. Petrichenko, and K. A. Safaryan, in *Ferrites*, Proc. International Conference on Ferrites, Tokyo (1992).
[12] V. A. Bokov, V. V. Volkov, and N. L. Petrichenko, Pis'ma Zh. Tekh. Fiz. **19**(11), 89 (1993) [Tech. Phys. Lett. **19**, 734 (1993)].
[13] a) V. M. Eleonskiĭ and N. E. Kulagin, Zh. Éksp. Teor. Fiz. **84**, 616 (1983) [Sov. Phys. JETP **57**, 356 (1983)]; Zh. Éksp. Teor. Fiz. **93**, 1436 (1987) [Sov. Phys. JETP **66**, 819 (1987)]; b) V. M. Eleonskiĭ, N. N. Kirova, and N. E. Kulagin, Zh. Éksp. Teor. Fiz. **77**, 409 (1979) [Sov. Phys. JETP **50**, 209 (1979)].
[14] E. Schlemann, Appl. Phys. Lett. **19**, 274 (1978); V. G. Bar'yakhtar, B. A. Ivanov, and A. L. Sukstanskiĭ, JETP Lett. **27**, 211 (1978); Fiz. Tverd. Tela (Leningrad) **20**, 2177 (1978) [Sov. Phys. Solid State **20**, 1257 (1978)].
[15] N. E. Kulagin and A. F. Popkov, JETP Lett. **43**, 249 (1986); Fiz. Met. Metalloved. **63**, 429 (1987).
[16] V. M. Eleonskiĭ, N. N. Kirova, and N. E. Kulagin, Zh. Éksp. Teor. Fiz. **76**, 705 (1979) [Sov. Phys. JETP **49**, 352 (1979)].
[17] V. G. Bar'yakhtar, Zh. Éksp. Teor. Fiz. **87**, 1501 (1984) [Sov. Phys. JETP **60**, 863 (1984)].
[18] V. G. Bar'yakhtar, Fiz. Tverd. Tela (Leningrad) **29**, 1317 (1987) [Sov. Phys. Solid State **29**, 754 (1987)].
[19] V. G. Bar'yakhtar, B. A. Ivanov, and K. A. Safaryan, Solid State Commun. **72**, 1117 (1989); V. G. Bar'yakhtar, V. A. Brodovoĭ, B. A. Ivanov, I. V. Krutsenko, and K. A. Safaryan, Fiz. Tverd. Tela (Leningrad) **32**, 852 (1990) [Sov. Phys. Solid State **32**, 502 (1990)].
[20] B. A. Ivanov and S. N. Lyakhimets, Zh. Éksp. Teor. Fiz. **100**, 901 (1991) [Sov. Phys. JETP **73**, 497 (1991)].
[21] E. G. Galkina, B. A. Ivanov, and K. A. Safaryan, Zh. Éksp. Teor. Fiz. **111**, 158 (1996) [JETP **84**, 87 (1996)].

Translated by Frank J. Crowne

Donors in a strong magnetic field and elastic magnetic-impurity resonance in diamondlike semiconductors

S. M. Dikman* and V. M. Zhilin

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

(Submitted 6 February 1997)
Zh. Éksp. Teor. Fiz. **112**, 975–1010 (September 1997)

The possibility of resonance during elastic intravalley scattering in n-type semiconductors is investigated in connection with the crossing (due to anisotropy of the effective mass) of the energy levels of excited states of a shallow donor as functions of the magnetic field. The hybridization of states of different frequencies in the vicinity of a crossing is attributed to the emergence of a nonzero dipole moment of the excited impurity atom and, accordingly, a long-range potential, which creates carrier-transport anomalies. The lower part of the donor spectrum is calculated as a function of the magnetic field in Si with $\mathbf{B}\|\langle 001\rangle$ and in Ge with $\mathbf{B}\|\langle 111\rangle$ or $\mathbf{B}\|\langle 110\rangle$. A crossing occurs in Ge in the field range 9.9 T$<B<$16.7 T and in Si in the field range 10.5 T$<B<$37.7 T. The characteristic longitudinal relaxation time and the transverse conductivity, which are determined by scattering at excited donors in the presence of the hybridization of states, are calculated. [S1063-7761(97)01509-6]

1. INTRODUCTION

We propose to investigate the eigenstates of a shallow impurity (donor) and the transport of charge carriers, which depends on scattering by the impurity, in a quantizing magnetic field. Scattering of this kind exhibits several prominent features of the magnetoresistance, which are classified under the general heading of magnetic-impurity resonances.[1,2] For the most part they are inelastic resonances, which appear when the energy difference between the excited and ground states of the impurity is a multiple of the cyclotron energy in the conduction band.[3–5] One such inelastic resonance is the exceptional scattering anomaly observed in tellurium in a magnetic field, where the first (due to intervalley splitting of the $1s$ state) and second (type $2p_-$ state) acceptor excitation levels are equidistant.[6] Finally, there is a phenomenon that has been predicted theoretically[7] but has yet to be observed: elastic intervalley scattering by the impurity potential in magnetic fields corresponding to the condition that the energies in two nonequivalent valleys coincide. We emphasize that the principal objective of the present study is to analyze theoretically an altogether new resonance, the possible existence of which has been suggested earlier.[8,9] Such a resonance is already present in intravalley scattering in bulk n-type semiconductors and is associated with the crossing (by virtue of anisotropy of the effective mass) of $2s$-like and $2p_-$-like energy levels E_{2s} and E_{2p_-} of donor excited states regarded as functions of the magnetic field. The fields corresponding to the crossings are special points of magnetic degeneracy. This situation, in turn, produces a nonzero dipole moment of the excited impurity atom and, as a result, a long-range potential that decays as $1/r^2$ and leads to anomalies in the scattering of electrons. The investigation of the phenomenon primarily subsumes two problems: first, to determine the magnetic fields corresponding to the resonance and, second, to study the behavior of the low-temperature mobility in the vicinity of such an anomaly.

We know that the first problem reduces to the calculation of the states of the shallow impurity in the nondegenerate zone in a strong magnetic field and is obviously of independent interest. It has been investigated in a number of papers both experimentally (in Refs. 10–12 for Si and in Ref. 13 for Ge and GaAs) and theoretically (by perturbation theory for Si in Ref. 14 and by the application of a variational method to Ge in Refs. 15 and 16). It is interesting that the theory in this case reduces to a combination of two well-known problems of single-particle quantum mechanics: the problem of the hydrogen atom in a strong magnetic field (the so-called quadratic Zeeman effect[17–29]) and the problem of a donor with uniaxial anisotropy of the effective mass.[30–34] In both cases the Schrödinger equation is two-dimensional by virtue of axial symmetry, but it cannot be solved analytically. Of little use are asymptotic expressions obtained in the limiting cases of large and small dimensionless parameters characterizing the strength of the magnetic field or the anisotropy. For example, the parameter in the quadratic Zeeman effect is the magnetic field reduced to dimensionless form by the Elliott–Loudon criterion:[23] $\mathscr{B}=(\hbar^3\kappa^2/e^3m^2c)|\mathbf{B}|$ ($\mathbf{B}$ is the field in standard units, κ is the static permittivity, and m is the effective mass). The perturbation series in $\mathscr{B}^2$ yields a satisfactory result in the determination of the state wave functions only for $\mathscr{B}<0.1-0.15$ (Refs. 14 and 22), and the asymptotic solution in strong fields[23] is valid for a large value of $\ln\mathscr{B}$, which is experimentally unrealistic, at least for Si, Ge, and GaAs. A similar situation is encountered in the ratio $\gamma=m_\perp/m_\|$ characterizing the anisotropy. For small values of γ the true parameter is actually $\gamma^{1/3}$ (see Ref. 30), and even for germanium (in which $\gamma=0.05134$) the asymptotic result obtained in the limit $\gamma\to 0$ gives a large error.

The original approach to the approximate solution of these problems was a variational method (see, e.g., Refs. 15–21, 30–32), which yields a good result for the energy levels $E_i(B)$. As a rule, however, despite the ease of visualization of the final result, the method is not suitable for finding the state wave functions. This fact has been demonstrated conclusively by Turbiner[21] in a comparison of the quadrupole moment Q_{zz} of the hydrogen atom in the ground state in a field $\mathscr{B}=0.1$, determined from perturbation theory, with the corresponding values calculated on various types of trial wave functions determined in the course of solving the variational problem. The result for Q_{zz} is found to be extremely (to the extent of a difference in order of magnitude) dependent on the type of trial wave functions chosen, even though the calculated ground-state energy is scarcely sensitive to the form of the trial functions. In strong fields, beginning approximately with $\mathscr{B}\gtrsim 10$, a more accurate approach is the so-called adiabatic method of approximate separation of variables used some time ago in Ref. 23,[1)] (see also Ref. 24), but the value of $\mathscr{B}$ obtained experimentally in a semiconductor, as a rule, is considerably lower.

It is important to mention so-called exact methods, which are used extensively, especially in recent times. The concept involves two types of solutions. The first type is a numerical solution of the problem, which provides, in principle (on an ''ideal'' computer), a means for finding the energy levels and wave functions with arbitrarily high accuracy. Solutions of this kind have been obtained for the hydrogen atom in a magnetic field[25–28] and for a donor in zero field.[33] The second type of solution is a realistic analytical solution of the problem. The result obtained for a radial wave function in this case is represented by a series in r, in which the coefficients of the expansion are calculated analytically by a recursive procedure. A fairly general description of this method is given in Ref. 34, along with an illustration of its application to the donor problem. In application to the hydrogen atom, an exact solution of the second type has been obtained very recently[29] with the attainment of an unprecedented number of significant figures in the energy spectrum computations. The second type of solution actually implies a dramatic reduction in the computing time from Refs. 25–28, and 33. The accuracy of all these calculations far surpasses the accuracy of the effective-mass method itself, which is the method used in constructing the band structure model for semiconductors, and is excessive for our purposes. Moreover, the state wave functions are determined in a very cumbersome form and are unsuitable for subsequent calculations. Finally, it must be borne in mind that the principal axis of the effective-mass ellipsoid in our case, generally speaking, is not in the same direction as the magnetic field. The Schrödinger equation, in principle, becomes three-dimensional, so that an exact solution of either the first or the second type will scarcely be feasible for the combined problem in the near future.

The sum total of these considerations has convinced us of the need to return to a variational method in solving the spectral problem. A key issue is the choice of the type of trial wave functions (Sec. 2). We draw on the ideas of Refs. 20 and 21, so that the form of the functions is largely determined by the symmetry of the Schrödinger equation and the asymptotic properties of the sought-after solutions. In the final analysis, this approach in any case affords the possibility of calculating the matrix elements of the dipole and quadrupole moments (as verified by several direct and indirect tests) within error limits fully compatible with the limits of the effective-mass approximation. The results of the calculations are given for Si and Ge in Sec. 3, based on the two-dimensional Schrödinger equation (effective-mass ellipsoid oriented along the field **B**), and in Sec. 4 for the three-dimensional problem (**B** perpendicular to the principal axis of the valley).

Finally, Sec. 5 is devoted to the second of the above-mentioned problems, i.e., the magnetic-impurity resonance problem *per se*. Here we calculate the principal contribution to the longitudinal and transverse impurity conductivities near the point of intersection of the levels for silicon with $\mathbf{B}\|\langle 001\rangle$ and for germanium with $\mathbf{B}\|\langle 111\rangle$ $\mathbf{B}\|\langle 110\rangle$. We rely on the well-known theory[2,35] first formulated decades ago.[36–38] However, it is impossible to make direct use of published results for the point and Coulomb potentials, because in our case the conductivity is governed by scattering at an uncharged impurity, which nonetheless exerts a long-range influence. We consider the case in which the main contribution is from small-angle carrier scattering at distances much greater than the effective Bohr radius of the donor. The interaction potential can be assumed to depend only on the free-carrier coordinates, and phenomena associated with the three-body problem[39] can be disregarded. Results are obtained in final form for Si in a field **B** directed along $\langle 001\rangle$.

In Sec. 6 we discuss possible experimental verification techniques. We call attention to the fact that resonance is mentioned strictly in the terminological sense. If $\mathbf{B}\|z$ holds and the electric field satisfies $\mathbf{F}\|y$, the experimental magnetoresistance must be expected to exhibit nonmonotonic behavior in the graphs of $\rho_{zz}(\mathbf{B})$ and $d\rho_{yy}(\mathbf{B})/dB$ near the corresponding value $\mathbf{B}=\mathbf{B}_0$. A certain complexity is also created by the need to ensure the existence of free carriers simultaneously with neutral (but excited!) donors in the sample. All the same, we believe that the predicted phenomenon is fully observable at least for such a material as germanium. Also in Sec. VI, on the basis of the now known donor spectrum, we predict the possibility of other types of resonances and indicate the expected resonance values of the magnetic field.

Some of the results pertaining to the solution of the spectral problem for a donor in Si have been published previously.[8,9] We give them in Secs. 3 and 4 without any further citation, along with calculations of the wave functions and energy levels of other excited states.

2. CHOOSING THE TRIAL FUNCTIONS; DESCRIPTION AND JUSTIFICATION OF THE METHOD

We intend to solve the Schrödinger equation in a uniform field **B** with the effective-mass Hamiltonian for a nondegenerate energy band. We consider only the cases of highest symmetry, where the magnetic field is strictly parallel or

strictly perpendicular to the axis of revolution of the effective-mass ellipsoid, with more detailed attention given to the first configuration.

2.1. Magnetic field aligned with the axis of revolution of the effective-mass ellipsoid (two-dimensional problem)

The substitution $z=z'\sqrt{\gamma}$ and the normalization of quantities to dimensionless form reduces the Schrödinger equation to

$$\left(-\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}-\frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2}-\frac{\partial^2}{\partial z'^2}-i\mathscr{B}\frac{\partial}{\partial\varphi}+\frac{1}{4}\mathscr{B}^2\rho^2-\frac{2}{\sqrt{\gamma z'^2+\rho^2}}\right)\Psi=\mathscr{E}\Psi. \tag{2.1}$$

This process transfers the anisotropy into the Coulomb term. The dimensionless quantities are referred to the following units of length, energy, and magnetic field:[2)]

$$\begin{aligned} &a_B=\frac{\hbar^2\kappa}{m_\perp e^2}\simeq(99.9,\,31.7)\ \ \text{Å},\\ &\frac{m_\perp e^4}{2\hbar^2\kappa^2}\simeq(4.68,\,19.94)\ \ \text{meV},\\ &\frac{e^3m_\perp^2c}{\hbar^3\kappa^2}\simeq(6.59,\,65.6)\ \ \text{T}. \end{aligned} \tag{2.2}$$

The first number in each pair applies to germanium, and the second number to silicon. Equation (2.1) is written in cylindrical coordinates and is in fact two-dimensional, since the dependence on the angle φ is characterized by the quantum number m ($\Psi\propto e^{im\varphi}$). To justify the choice of trial wave function, all the same, it is advisable to use spherical coordinates as well: $r=\sqrt{\rho^2+z'^2}$, $\sin\theta=\rho/r$, $\cos\varphi=x/\rho$.

Of first priority is the trial wave function of the ground state. We substitute $\Psi=e^{-\Phi(\mathbf{r})}$ into Eq. (2.1) and investigate the solution of the equation for Φ in the asymptotic region of large r (for fixed angles θ and φ) or large $\mathscr{B}$. To find the principal terms of the expansion of Φ in powers of r, we can disregard the Coulomb potential in Eq. (2.1). As a result, it is readily established that the principal term of the expansion is the squared term and is uniquely determined irrespective of the form of the attraction potential, while the term next in order, i.e., the linear term, is determined to within a constant (independent of the angles θ and φ) factor evaluated after matching with the solution for finite r:

$$\Phi|_{r\to\infty}=\frac{1}{4}\rho^2|\mathscr{B}|+C|z'|. \tag{2.3}$$

It is fairly obvious that it would be incorrect to use the terms of the expansion (2.3) directly for the ''construction'' of a trial wave function describing the solution for any **r** and any **B**, because the first term has an artificial singularity at zero for **B**, and the second term has the same for z'. This difficulty is circumvented if the role of the trial wave function is assigned, not to the expansion of Φ in r, but to the expansion of the square of Φ with two principal terms in the form

$$\Phi^2\sim\frac{\mathscr{B}^2\rho^4}{16}+\sqrt{1+\mathscr{B}^2}(u\rho^2+wz'^2)\sqrt{z'^2+v^2\rho^2}. \tag{2.4}$$

In this expression u, w, and v automatically emerge as the parameters to be varied and must obey the following relation asymptotically:

$$v|_{\mathscr{B}\to\infty}=w/u|_{\mathscr{B}\to\infty}=0. \tag{2.5}$$

The term next in order after those included in (2.4) is the squared term in r. On the other hand, it is the principal term in the expansion of Φ^2 in the vicinity of $r=0$ and determines the linear term in r in the expansion of Φ for the solution of the Schrödinger equation in zero field $\mathscr{B}$. For $\gamma\neq0$ the coefficient of the latter term depends on the angle θ, and this dependence, being described by an ordinary differential equation in the variable θ, can be found in principle. Such an approach, however, would be impractical and in application to the variational method would represent overkill in terms of accuracy. On the other hand, it has been known since the fifties[30] that the anisotropy of the effective mass can be taken into account with complete success by choosing the trial wave function Φ in the form $\sqrt{a^2(\rho^2+b^2z'^2)}$ with the a and b as the variational parameters. As a result, we conclude with the recommendation that the trial wave function of the ground state can be modeled over the entire ranges of r and **B** by $\Phi(\mathbf{r})$ in the form

$$\Phi^2=a^2(\rho^2+b^2z'^2)+\sqrt{1+\mathscr{B}^2}(u\rho^2+wz'^2)\times\sqrt{z'^2+v^2\rho^2}+\frac{\mathscr{B}^2\rho^4}{16}. \tag{2.6}$$

We now turn our attention to the symmetry properties of the desired solutions of the Schrödinger equation. The symmetry group of the effective-mass Hamiltonian is $D_{\infty d}$.[3)] A unitary representation of this group is used to transform the function (2.6) and, eventually, the ground-state wave function.

We seek other solutions of the Schrödinger equation in the form

$$\Psi_i=R_ie^{-\Phi_i}, \tag{2.7}$$

assuming that the exponent has a form analogous to (2.6). Each ith solution of Eq. (2.7), of course, is determined by its own set a_i, b_i, u_i,... of variational parameters of the function Φ_i. The symmetry of the solution is determined entirely by the function $R_i(\mathbf{r})$. We also require that the function R_i increase at most by a power law for large r. Specifically, the following relation must hold for any fixed asymmetric direction specified by the angles θ and φ:

$$R_i|_{r\to\infty}\propto r^{N-1},\quad N\geqslant1, \tag{2.8}$$

where N is a positive integer, which can be interpreted as the principal quantum number of the state. The complete set of all quantum numbers identifying the solution of the Schrödinger equation is described by the index i. We recall that the azimuthal number m is already known. It is readily shown that the restriction $|m|\leqslant N-1$ follows from condition (2.8).

Another characteristic of a state is the representation by which the function R_i is transformed. The form of this representation is obviously not in any way connected to N, but in general depends on m. In particular, if $m\neq 0$, the representation corresponding to R_i cannot be unitary. We consider only the lower part of the spectrum, i.e., states whose energies are below the zeroth Landau level (in any case for fields $\mathscr{B}$ lower than a certain finite value) and which go over to states of the discrete part of the Coulomb spectrum in the limits $\mathscr{B}\to 0$ and $\gamma\to 1$. We can therefore classify the investigated solutions of Eq. (2.1), retaining the notation of the corresponding Coulomb state. Letter indices govern the representation for which R_i is the basis function. The ground-state index i is obviously $1s$. We also look at other solutions of the Schrödinger equation (2.1) with $i=2s, 2p_m, 3p_m$ $(m=0,\pm 1)$, $3d_{\pm 1,\pm 2}$. With these considerations in mind we choose the function R_i in the form

$$\begin{aligned} &R_{1s}=1,\ R_{2s}=Q_{2s},\ R_{2p_0}=z',\ R_{2p_{\pm 1}}=\rho e^{\pm i\varphi},\\ &R_{3p_0}=z'Q_{3p_0},\ R_{3p_{\pm 1}}=\rho e^{\pm i\varphi}Q_{3p_{-1}},\ R_{3d_{\pm 1}}=z'\rho e^{\pm\varphi},\\ &R_{3d_{\pm 2}}=\rho^2 e^{\pm 2i\varphi}, \end{aligned} \tag{2.9}$$

where

$$Q_i=1-c_i\sqrt{m_i^2+\sqrt{z'^2+d_i^2\rho^2}+s_i^2(\rho^2+t_i^2z'^2)}. \tag{2.10}$$

All that requires explanation, obviously, is the form of the functions (2.10) governing the preexponential factor for the $2s$, $3p_0$, and $3p_{\pm 1}$ states. Thus, for the $2s$ state the condition (2.8) can also be satisfied by the function $1-c\sqrt{z'^2+d^2p^2}$ in the role of R_{2s}, but then in the range $r\lesssim 1$ (it is primarily in this neighborhood that the preexponential function is significant) we confine the discussion to the two principal, first- and second-order, terms in r for the preexponential factor of the trial wave function. We find that the inclusion of the next, squared term greatly improves the accuracy of the variational calculation. On the other hand, the simple addition of this term to the preexponential factor would violate the rule (2.8). In the final analysis Eq. (2.10) represents the most natural alternative for taking both requirements into account. The case $\mathscr{B}=0$, $\gamma=1$ corresponds, for example, to the following exact values of the coefficients in Eq. (2.10): $d_i=t_i=m_i=1$, $s_i=1/2$, $c_{2s}=1/2$, $c_{3p_0}=c_{3p_{-1}}=1/4$. We call attention to the fact that the dimensionless energies in states differing only in the sign of the quantum number m differ by $2|m|$; accordingly, the wave functions of these states have the same dependence on ρ and z'.

It is evident from Eqs. (2.7), (2.9), and (2.10) that the functions (Ψ_{1s},Ψ_{2s}), $(\Psi_{2p_0},\Psi_{3p_0})$, and $(\Psi_{2p_{-1}},\Psi_{3p_{-1}})$ are pairwise nonorthogonal. The corresponding states are orthogonalized automatically in the course of solving the problem. Here the general analytical scheme corresponds to that used in Refs. 15, 16, and 43 and entails the following. We assign the index 1 to the first state of the pair in equation, and assign the index 2 to the second state in the pair. Varying the parameters a_1, $b_1,\ldots$, primarily, we find the optimum wave function Ψ_1 minimizing the functional

$$\mathscr{E}(\Psi_1)=\frac{\langle\Psi_1|\mathscr{H}|\Psi_1\rangle}{\langle\Psi_1|\Psi_1\rangle}. \tag{2.11}$$

Here $\mathscr{H}$ is the dimensionless Hamiltonian [the operator acting on Ψ on the left-hand side of Eq. (2.1)]. From the known function Ψ_1 we then transform at once to the different basis

$$\begin{aligned} \Psi_1'&=C_{11}\Psi_1+C_{12}\Psi_2,\\ \Psi_2'&=C_{21}\Psi_1+C_{22}\Psi_2, \end{aligned} \tag{2.12}$$

where the coefficients C_{ij} must be determined by two requirements: 1) $\langle\Psi_1'|\Psi_2'\rangle=0$; 2) the quantities $\mathscr{E}_i\langle\Psi_i'|\mathscr{H}|\Psi_i'\rangle/\langle\Psi_i'|\Psi_i'\rangle$, treated as functions of C_{ij} for given Ψ_1 and Ψ_2, must each be a minimum.

As a result, the search for C_{ij} is the simplest application of the Ritz method (see Ref. 44 and the Appendix in Ref. 43) and reduces to the determination of the roots $\mathscr{E}_{1,2}$ of the equation $\det|H-\mathscr{E}Y|$, where H and Y are matrices with elements $H_{ij}=\langle\Psi_i|\mathscr{H}|\Psi_j\rangle$ and $Y_{ij}=\langle\Psi_i|\Psi_j\rangle$. The required coefficients are found as eigenvectors $\mathbf{C}_i=(C_{i1},C_{i2})$ corresponding to the nontrivial solution of the homogeneous linear system $(H-\mathscr{E}_iY)\mathbf{C}_i=0$. Finally, the larger of these two roots, specifically $\mathscr{E}_2$, is regarded as a functional on the space of functions Ψ_2. Varying the parameters $a_2,b_2,\ldots,c_2,m_2,\ldots$, we find $\min(\mathscr{E}_2)$, which gives us the energy in state 2, and the corresponding value of the root $\mathscr{E}_1$ gives the energy in state 1. The latter quantity is therefore a ''corrected'' version of the value (2.11) already found in the first stage of the solution. The normalized functions $\Psi_{1,2}'$ are treated below as eigenstate wave functions. For the orthogonal $3d_{-1,-2}$ states, naturally, the calculations are restricted to the first stage, i.e., minimization of the functional (2.11).

We call attention to a very important consideration, which can be exploited to indirectly estimate the computational error. The transformation (2.12) from the functions $\Psi_{1,2}$ to the functions $\Psi_{1,2}'$ formally violates condition (2.8) for state 1 only if the coefficient C_{12} has a nonzero value. This is because the function Ψ_2 found by the above-described method always contains the exponential function $\exp(-\Phi_2)$, which decays far more slowly than $\exp(-\Phi_1)$.[4)] In all the calculations, however, the coefficient C_{12} for the given pairs of states is very small in all the calculations, so that the correction introduced by it within our error limits has no effect whatsoever on the energy $\mathscr{E}_1$. The form of the function Ψ_1 is therefore an auspicious choice. In other words, condition (2.8) can be regarded as fulfilled for our functions within error limits consistent with the accuracy of the calculations involved. In the calculation of the matrix elements, on the other hand, the correction associated with the coefficient C_{12} becomes appreciable. Specifically, in calculating the quadrupole moment $Q_{zz}=\langle i|2z^2-\rho^2|i\rangle$ in the $i=1s$, $2p_0$, $2p_{-1}$ states, we have found that the transformation $\Psi_1\to\Psi_1'$ changes the value of Q_{zz} somewhat. The change, which is less than 1% over the entire range of $\mathscr{B}$, characterizes the error with which we can find the matrix elements in the given situation. Indeed, strictly speaking, in transforming to the optimal basis (2.12), it would be more proper, instead of the pair (1,2), to consider, say, the triple of states (1,2,3), in which state 3 has the same symmetry as 1

and 2 (for the investigated pairs of functions the three states are $3s$, $4p_0$, and $4p_{-1}$, respectively). Generally speaking, a correction to Ψ_1 of the form $C_{13}\Psi_3$ will be of the same order as $C_{12}\Psi_2$, but we drop it. Consequently, the difference in the results obtained by means of Ψ_1 and Ψ_1' can be used to estimate the error of the method. To include the term $C_{12}\Psi_2$, which characterizes the mixing of derivatives in the calculations by means of the wave function of state 1, is not necessary, because it is of higher order. Similarly, if the form of the function Ψ_2 is correctly ''divined,'' a correction to Ψ_2' of the form $C_{23}\Psi_3$ should be expected not to affect the results, at least not within the same error limits of the method.

The situation is altogether different in regard to the mixing $C_{21}\Psi_1$ in the second line of (2.12). Its inclusion does not violate condition (2.8), and so there is no reason to expect C_{21} to be small. Calculations show that for arbitrary γ and $\mathscr{B}$ the coefficient C_{21} is of the same order as C_{22}, and in all calculations involving the wave function of state 2, of course, Ψ_2' must be used in the form (2.12).

In addition, there is an abundance of direct and indirect clues to the effect that the wave functions found in our method are suitable, within fully acceptable error limits, for calculating both diagonal and off-diagonal matrix elements of the dipole and quadrupole moments. Following are some of these clues. First of all, in all the calculations we find that condition (2.8) applied to the asymptotic form always produces a wave function such that the equation $\Psi_{1,2}'(\mathbf{r})=0$ has $N-1$ real roots in r for fixed angles θ and φ. Although we lack rigorous proof of the necessity of this condition, it does seem natural and obviously correlates with the oscillation theorem for the one-dimensional Schrödinger equation.

An important indirect clue to the validity of the choice of trial wave function is the fact that the variational parameters governing the optimal trial function are almost always found to be weak, smooth functions of $\mathscr{B}$ for the investigated range of $\mathscr{B}$. Some of them, a_{1s} and b_{1s} for example, vary so slightly that their values corresponding to $\mathscr{B}=0$ can be fixed and regarded as constants up to $\mathscr{B}\sim 10$ without any significant loss of computational accuracy. For the two-dimensional problem the only exception from the general pattern is the parameter b for the d-states in germanium, which in the optimal trial wave function begins to drop abruptly at $\mathscr{B}>0.5$ and actually decreases to zero. This behavior, however, does not imply a drastic variation of the trial function itself, because its dependence on z, beginning with the same field values, is dictated by the parameter w, which increases by roughly an order of magnitude in comparison with its zero-field value. In principle, an appreciable reduction in the number of variational parameters is admissible in our method, but we forgo this operation, first of all, because it is impossible to state *a priori* which subset of the set of variable parameters involved in (2.9) and (2.10) is the best choice and, second, because observing the variations of the parameters governing the optimal trial wave function as the quantity $\mathscr{B}$ is varied provides an extra check on the validity of the calculations. In particular, the attainment of the asymptotic limit (2.5) has been verified directly for $\gamma=1$.

Direct comparison of our calculations with results obtained as part of the exact solutions of the two-dimensional problem[25–29,32,33] are made in the next section. Here we illustrate only one result of direct verification of the ground-state wave function for the case $\gamma=1$, $\mathscr{B}=0.1$. The quadrupole moment Q_{zz} can be found by perturbation theory and, to three-place accuracy (the second order of perturbation theory is used, i.e., terms $\sim\mathscr{B}^4$ are taken into account), is equal to 0.0231. The results obtained after variational calculations using various types of trial wave functions are given in Ref. 21; their scatter is more than an order of magnitude, with Q_{zz} ranging from 0.2327 to 0.0127 and, in this sense, the ''best'' of the trial functions in Ref. 21 corresponding to $Q_{zz}=0.0236$. On the other hand, the calculation of $Q_{zz}=\langle 1s|2z^2-\rho^2|1s\rangle/\langle 1s|1s\rangle$ using our function obtained from the minimization of the functional (2.11) (with $a_{1s}=0.9990$, $b_{1s}=1.0002$, $u_{1s}=1.065\times10^{-3}$, $v_{1s}=3.129$, and $w_{1s}=1.291\times10^{-3}$) yields a value that agrees with the perturbation-theoretic result out to three significant figures.

2.2. Magnetic field B perpendicular to the axis of revolution of the effective-mass ellipsoid (three-dimensional problem)

Let the magnetic field be directed along the z axis as before, and let the principal axis of the effective-mass ellipsoid be directed along the y axis. The symmetry group of the Hamiltonian with allowance for the variation of the vector **B** under spatial transformations is D_{2h}. After reduction to dimensionless form according to (2.2) and the substitution $\mathscr{B}=\mathscr{B}'/\sqrt{\gamma}$, $y=y'\sqrt{\gamma}$ we obtain the Schrödinger equation in the form

$$\left[-\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial y'^2}-\frac{\partial^2}{\partial z^2}+i\mathscr{B}'\left(y'\frac{\partial}{\partial x}-x\frac{\partial}{\partial y'}\right)+\frac{1}{4}\mathscr{B}'^2(x^2+y'^2)-\frac{2}{\sqrt{\gamma y'^2+x^2+z^2}}\right]\Psi=\mathscr{E}\Psi. \tag{2.13}$$

The guidelines for choosing the trial wave function are essentially the same as those used in the two-dimensional problem. We write these functions in the form (2.7), requiring that Φ_i and R_i satisfy conditions (2.3) and (2.8), but now $\rho=\sqrt{x^2+y'^2}$ and $r=\sqrt{\rho^2+z^2}$. The result is a natural generalization of Eqs. (2.9) and (2.10). We consider the six states for which

$$\Phi_i^2=a_i^2(z^2+x^2+b_i^2y'^2)+\sqrt{1+\mathscr{B}'^2}(u_ix^2+k_iy'^2+w_iz^2)\sqrt{z^2+v_i^2(x^2+t_i^2y'^2)}+\mathscr{B}'^2(x^2+y'^2)^2/16, \tag{2.14}$$

$$R_{1s}=1,\ R_{2s}=Q_{2s},\ R_{2p_0}=hx-iy',\ R_{2p_-}=x+igy',$$

$$R_{2p_-}=z,\ R_{3p_-}=zQ_{3p_-}, \tag{2.15}$$

$$Q_i=1-c_i\times\sqrt{m_i^2+\sqrt{z^2+(l_ix)^2+(d_iy')^2}+s_i^2[z^2+n_i^2(x^2+q_i^2y'^2)]}. \tag{2.16}$$

Note that we keep the same notation as that used in Refs. 10–13, i.e., the indices correspond to the indices of axisymmetric states in zero field. If we make the transition to the axisymmetric problem, removing anisotropy ($\gamma \to 1$), the nonorthogonal states (2.14)–(2.16), $2p_0$ and $2p_+$, go over to the states $2p_{-1}$ and $2p_{+1}$ of the hydrogen atom in a magnetic field (with $g=h=1$), while the $2p_-$ and $3p_-$ go over to $2p_0$ and $3p_0$, respectively. As $\mathscr{B}' \to 0$, the states described by the functions $\Psi_{2p_\pm}$ must, each independently, go over to the twofold-degenerate $2p_{\pm 1}$ state of the two-dimensional problem in zero field (where, as a result, the solutions of the variational problem must be $g \to 0$).

The pairs (Ψ_{1s}, Ψ_{2s}), $(\Psi_{2p_0}, \Psi_{2p_+})$, and $(\Psi_{2p_-}, \Psi_{3p_-})$ in the three-dimensional problem are orthogonalized by the same scheme as described above. For the $(1s,2s)$ and $(2p_-, 3p_-)$ states the mixing coefficient C_{12} in Eq. (2.12) is again small (of the order of 10^{-4}–10^{-3} times C_{11}) and does not affect the energy of the state. A comparison of the values of the quadrupole moment calculated by means of the functions Ψ_{1s} and Ψ'_{1s} or Ψ_{2p_-} and Ψ'_{2p_-} makes it possible, as in the two-dimensional problem, to estimate the error made in calculating the matrix elements. These estimates show that the resulting wave functions can be used in the three-dimensional case to obtain results for the dipole and quadripole moments within 10%. For the $2p_0$ and $2p_+$ states, on the other hand, the orthogonalization (2.12) does not violate condition (2.8), so that no basis exists for expecting the coefficient C_{12} to be small in this case. However, a calculation shows that $|C_{12}|$ is not more than a few percent of $|C_{11}|$, and the corresponding correction to the energy $\mathscr{E}_1$ is on the order of one tenth of one percent.

3. DONOR LEVELS FOR VALLEYS ORIENTED ALONG THE FIELD B

The case discussed in this section is encountered in germanium if $\mathbf{B} \| \langle 111 \rangle$ and in silicon if $\mathbf{B} \| \langle 001 \rangle$. The issue here is ellipsoids of the constant-energy surface which, in the terminology of Refs. 13, 15, and 16, refer to case A. In light of the many-valley property the corresponding states in germanium are not degenerate, and in silicon they are twofold degenerate. In a magnetic field the energy levels of these states are higher than the levels of the triplet in Ge [three ellipsoids of so-called type B (Refs. 13, 15, and 16) oriented at a $70°32'$ angle relative to the direction of $\mathbf{B}$] and the quadruplet in Si (four valleys of type E; see the next section).

In this article we stay within the bounds of the effective-mass approximation and, hence, completely ignore the chemical shift of the levels and their valley-orbital splitting that occurs when the potential of the central cell is taken into account (see, e.g., Refs. 45 and 33), i.e., the short-range component of the (as a rule, completely unknown) impurity-ion field. In reality, intervalley effects are significant only for the $1s$ state, whose energy level E_{1s} does not cross energy levels of other states of the same valley. This implies, in any case, that the $1s$ state never participates in the investigated resonance, so that the value of $\mathscr{E}_{1s}(\mathscr{B})$ and the corresponding wave function are of not importance to us. On the other hand, the type of intervalley splitting for a substitutional impurity (which a shallow donor usually is) can be easily described qualitatively by means of a group-theoretic analysis analogous to that employed for $\mathbf{B}=0$ (Ref. 45). An example of such an analysis is given in Appendix 1 in application to S-states for the magnetic field directions considered in this article. The only anomaly lies in the fact that the point group of directions of a crystal with an imbedded magnetic field is now described by the corresponding magnetic class.

It should be noted that the most systematic quantitative calculation of the wave functions with the central-cell correction is given in Ref. 33 for the case $\mathbf{B}=0$. A more primitive calculation of valley–orbital splitting, but now with allowance for a finite field $\mathbf{B}$, has been carried out for a Ge crystal in the same paper, where even in zero field the splitting of the $2s$ level is found to far exceed the experimental value (cf. Ref. 13 in this regard). Also, an explicit error is incurred in Ref. 15 for the case $\mathbf{B} \| \langle 001 \rangle$, resulting in quantitatively incorrect quadruplet splitting (see Appendix 1).

First of all, we illustrate zero-field results and compare them with exact calculations based on the effective-mass approximation[33] and with other calculations.[31,32] All the results are summarized in Table I. We note in this connection that the method used in Ref. 32, although called variational, is really the Ritz method in essence, where the varied parameters are the coefficients of the expansion of the unknown radial wave function into a fixed system of simple basis functions. If the dimensionality of such a basis is not restricted in some way, a result is obtainable with arbitrarily high accuracy. In the calculation of the energies of even states in Ref. 32 the number of basis functions reaches 105. In other words, this variational method involves 105 fitting parameters. Clearly, to use a wave function in this form in subsequent calculations would be impractical in the extreme.

Table I does not include any experimental data for comparison. Such measurements for common semiconductor materials in zero magnetic field have been published some time ago (see, e.g., the survey in Ref. 46, as well as Ref. 47), and they have been analyzed in detail.[32,33,41] As a result, it is safe to say that within error limits lower than those of spectroscopic measurements the energy level calculations in the effective-mass approximation agree with the experimental values for all excited states except $2s$, $3s$, and $3d_0$, but the chemical shift for these three levels does not amount to more than a few percent of the energy itself. A more important consideration from our standpoint is to compare the matrix elements found by means of our wave functions with the ''exact'' wave functions. The oscillator strengths of the optical transitions to P-states from the $1s$ ground state have been calculated.[33] The values of the corresponding dimensionless quantities

$$\frac{\gamma|\mathscr{E}_{Np_0}-\mathscr{E}_{1s}|}{2+\gamma}\langle 1s|z'|Np_0\rangle^2 \quad \text{and} \quad \frac{|\mathscr{E}_{Np_{\pm 1}}-\mathscr{E}_{1s}|}{2+\gamma} \times \langle 1s|\rho e^{\mp i\varphi}|Np_{\pm 1}\rangle^2, \tag{3.1}$$

obtained in Ref. 33 (in the effective-mass approximation) and calculated by us are also given in Table I. The angle brackets in Eq. (3.1) signify averaging in the dimensionless basis x,y,z (but not x,y,z'!). The states are assumed to be

TABLE I. Zero magnetic field.

State		Binding energy, meV			Oscillator strength (3.1), 10^{-3}	
	i	Our calculations	Ref. 33	Ref. 32	Ref. 31	Our calculations	Ref. 33
Ge		$\|E_i\|=4.676\|\mathscr{E}_i\|$ [meV]					
	$1s$	9.78 (9.73)	9.78	–	9.78	–	–
	$2p_0$	4.75 (4.71)	4.75	4.78	4.74	17.9	18.8
	$2s$	3.58 (3.49)	–	3.60	3.52	–	–
	$3p_0$	2.57 (2.54)	2.57	2.59	2.56	2.29	1.91
	$2p_{\pm1}$	1.72 (1.72)	1.72	1.73	1.73	233	233
	$3d_{\pm1}$	1.26	–	1.27	–	–	–
	$3p_{\pm1}$	1.04	1.04	1.04	1.03	42.5	40.6
	$3d_{\pm2}$	0.674	–	0.68	–	–	–
Si		$\|E_i\|=19.94\|\mathscr{E}_i\|$ [meV]					
	$1s$	31.25	31.27	–	31.27	–	–
	$2p_0$	11.48	11.49	11.49	11.51	56.1	57.9
	$2s$	8.85	–	8.86	8.83	–	–
	$2p_{\pm1}$	6.40	6.40	6.40	6.40	289	287
	$3p_0$	5.48	5.485	5.49	5.48	8.43	7.81
	$3d_{\pm1}$	3.87	–	3.87	–	–	–
	$3p_{\pm1}$	3.12	3.12	3.12	3.12	55.0	53.9
	$3d_{\pm2}$	2.63	–	2.63	–	–	–

Notes: Our calculations of the energies E_i of the states of the lower part of the spectrum of shallow donors in Ge and Si are compared with previous calculations[31–33] in the effective-mass approximation. The figures in parentheses give the values found in Refs. 15 and 16 for the case $\mathbf{B}=0$. The oscillator strengths (3.1) of the optical transitions $Np_m \rightarrow 1s$ are calculated for P-states. The next to the last column gives the corresponding dimensionless quantity (3.1), and the last column gives the same quantity determined from an ''exact'' calculation.[33]

normalized in the same basis. Clearly, the agreement with the exact calculation allowing for the above-mentioned peculiarities of the variational method is fully satisfactory. We note that the calculations of the matrix elements always give approximately the same absolute error, which is dictated by the error of the previously determined wave functions. The relative error increases if the unknown quantity in our dimensionless units is small.

Figures 1 and 2 shows the results of calculations of the spectra of the energy, referred to the zeroth Landau level, of donor excited states in germanium and silicon, respectively. The crossing of the $2s$ and $2p_{-1}$ levels at $B=9.9$ T in Ge and 10.5 T in Si is enclosed in the box. We call attention to the inversion of the levels of the $2p_{-1}$ and $3p_0$ states in weak fields (e.g., up to 2 T) in a Ge crystal. Data from measurements of the absorption lengths from the photoconductivity in Ge(Sb) (Ref. 13) are indicated in Fig. 1. A striking feature here is the fact that the transition energy between two states has always been measured in experiments, so that the dependence of the energy on B for one of the spectral levels must be normalized to the theoretical value. We have normalized both to the $2p_0$ level (dark circles) and to the $2p_{-1}$ level (light circles). As a result, it is evident that normalization to the $2p_{-1}$ level is preferable for determining the energy of the $2s$ state. The relative position of the $2p_0$ and $2p_{-1}$ levels is determined by adding the Zeeman components of the $2p_0 \rightarrow 2s$ and $2s \rightarrow 2p_{-1}$ transitions. Also, only if the $2s$ state is involved can the $3p_0$ level be determined from the measurement data in Ref. 13 in fields higher than 2.5 T ($2s \rightarrow 2p_{-1}$ and $2s \rightarrow 3p_0$ transitions), and only below 1 T is this level calculated by means of the difference in the components $2p_0 \rightarrow 3d_{-1}$ and $3p_0 \rightarrow 3d_{-1}$). On the other hand, the $2s$ state is stronger than the other excited states and is subject to the influence of effects beyond the scope of the effective-mass approximation, primarily effects such as the chemical shift and valley–orbital splitting, as well as broadening arising because the position of the impurity center does not coincide with a lattice site. As a result, the sum or difference of the measured components must differ significantly from the true difference in the energies $|E_{2p_{-1}}-E_{2p_0}|$ and $|E_{3p_0}-E_{2p_{-1}}|$. The effect should intensify as B is increased (as is indeed observed; see Fig. 1), because as the wave function becomes ''compressed,'' the role of the central cell in the formation of the effective potential increases, and the latter differs more and more from the Coulomb potential. In our opinion, this fact (not the computational error) is the source of the discrepancy with theory, specifically for the part of the experimental spectrum determined using the $2s$ level.

Experiments in silicon are always performed, of course, in much weaker dimensionless fields $\mathscr{B}$. For example, according to measurements of the spectrum of the photoexcitation $1s(A_1) \rightarrow Np_{\pm1}$ (A_1 is a unitary representation of the lower split doublet; see Fig. 7 in Appendix 1) in Si(P), the maximum value of $\mathscr{B}$ must be $\simeq 0.1$ (Ref. 10) and $\simeq 0.15$ (Refs. 11 and 12). Nonetheless, owing to the anisotropy of the effective mass, perturbation calculations are unsuitable even for determining the state energy (see Ref. 14), let alone the wave functions. The application of the Ritz variational method in the basis of eigenfunctions of Coulomb states helps to improve the agreement between calculations and

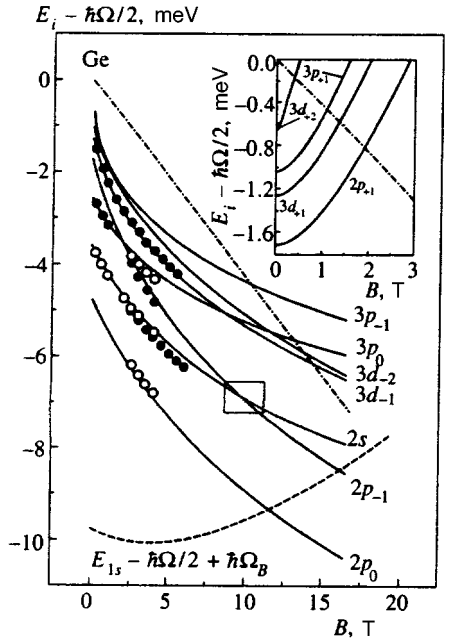

FIG. 1. Energy levels of shallow donors in germanium in a field $\mathbf{B}\|\langle 111\rangle$ for an A-type valley with axis of revolution parallel to $\mathbf{B}$. The energy is read from the zeroth Landau level of the same valley. Inset: Energies in states with a positive quantum number m (they are calculated by summing the energy of a level with $m<0$ with the quantity $|m|\hbar\Omega$). The edge of the conduction band (zeroth Landau level of three other B-valleys), referred to the zero level of the A-valley, is represented by a dot–dash line in the main figure, and in the inset it corresponds to the energy $\hbar(\Omega_B-\Omega)/2$, where $\Omega_B=\Omega\sqrt{1+8\gamma}/3\approx 0.396\Omega$. The dashed curve represents the energy of the $1s$ state shifted upward by the amount of the cyclotron energy of conduction electrons. Its points of intersection with the solid curves determine (in the effective-mass approximation) the positions of the principal inelastic magnetic-impurity resonance peaks. The point symbols represent experimental data;[13] the graphs are normalized to the $2p_{-1}$ (○) and $2p_0$ (●) levels.

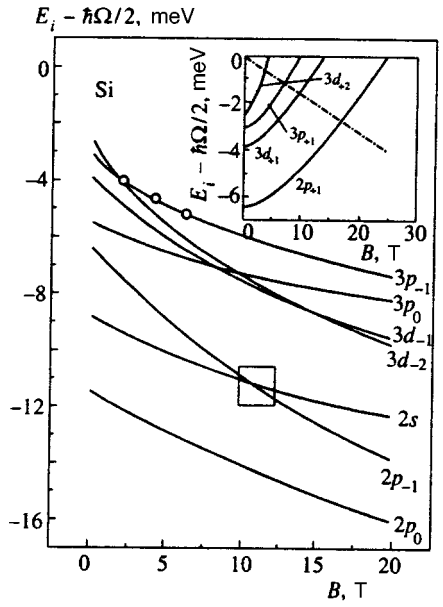

FIG. 2. Energy levels of shallow donors in silicon in a field $\mathbf{B}\|\langle 001\rangle$ for an A-type valley. The notation is the same as in Fig. 1. The edge of the conduction band (dot–dash line) is determined by E-valleys oriented along $\langle\pm 100\rangle$ and $\langle 0\pm 10\rangle$. In the figure it corresponds to the energy $\hbar(\Omega_E-\Omega)/2$, where $\Omega_E=\Omega\sqrt{\gamma}\approx 0.456\Omega$. The point symbols represent experimental data;[10] the graphs are normalized to the $2p_{-1}$ level.

experiment (see Ref. 11), but it is difficult to understand the graphically displayed experimental results in Refs. 11 and 12. The problem is that in these measurements the level differences $\Delta_{2p}=E_{2p_{+1}}-E_{2p_{-1}}$, $\Delta_{3p}=E_{3p_{+1}}-E_{3p_{-1}}$, and $\Delta_{4p}=E_{4p_{+1}}-E_{4p_{-1}}$ for the type-A effective-mass ellipsoid are far from equal (within error limits far exceeding the usual ones for spectroscopic measurements) to each other or to the quantity

$$\hbar\Omega=\hbar eB/m_\perp c.$$

We do not see any reasonable explanation for this, so that only the experimental data of Ref. 10 for $B=2$ T, 4 T, and 6 T are indicated in Fig. 2, and unfortunately even these are shown only graphically, but for them the equation $\Delta_{2p}(B)=\Delta_{3p}(B)=\hbar\Omega(B)$ is satisfied within error limits fully compatible with the measurement error. Based on the data of Ref. 10 for the type-A ellipsoid, only the energy $E_{3p_{-1}}$ can be found by normalizing to the $E_{3p_{-1}}$.

It is useful to compare our results with the corresponding values calculated in Refs. 15 and 16 by a method that differs from our own only in the choice of the trial wave function. The comparison is shown partially (for zero field) in Table I. For other values of B our calculations give approximately 1% higher or even lower values of $\mathscr{B}-\mathscr{E}_i$ than in Refs. 15 and 16. Such a disparity would be imperceptible in the graphs of Fig. 1. We recall that the important consideration here is not this somewhat minor refinement of the binding energy, but the capability of using our wave functions to calculate the matrix elements, which is not offered by the wave functions used in Refs. 15 and 16.

We shall not give the specific values of the parameters $a_i, b_i, \ldots$, obtained from the calculations, assuming that the calculations can be repeated with comparative ease by means of modern library programs. We interject one remark. Because of anisotropy the parameter u_i in the optimal trial wave function is found to be negative in weak fields; the effect is more pronounced for Ge, and it is altogether nonexistent if $\gamma=1$. On the other hand, in calculating $\langle\Psi_i|H|\Psi_j\rangle$ and $\langle\Psi_i|\Psi_j\rangle$, the ''cutoff'' of infinity in the numerical integration over ρ and z (the limit of integration is determined by the required accuracy) always takes place in the spatial domain where the exponent Φ_i is a monotonically increasing function of both ρ and z. This operation is possible because the parameter u_i is always fairly small in absolute value. For example, we have $u=-4.38\times 10^{-2}$ in a field $B=0$ for the $1s$ state in germanium, where $a=1.61$, $b=0.549$, $w=0.126$, and $v=2.74$. Out of all the negative values of u_i, here we give the largest in absolute value. As B increases, the optimal value of u_i increases, passes through zero approximately in the interval $\mathscr{B}=0.1-0.7$, and continues to increase monotonically, remaining positive. This behavior is characteristic of all the calculated states for both semiconductors; the only exception is the $3p_{-1}$ state in the sense that the value of u for Ge and for Si is positive from the very outset.

We conclude this section with a comparison of the results of our calculations for $\gamma=1$ in fields $\mathscr{B}=2$ and 20 with

TABLE II. Case $\gamma=1$.

	$1s$			$2p_{-1}$		
$\mathscr{B}$	$\mathscr{B}-\mathscr{E}_{1s}$	$\langle 1s\|r^2\|1s\rangle$	$\pi\|\Psi_{1s}(0)\|^2$	$\mathscr{B}-\mathscr{E}_{2p_{-1}}$	$\langle 2p_{-1}\|r^2\|2p_{-1}\rangle$	$\|\langle 1s\|\mathbf{r}\|2p_{-1}\rangle\|^2$
	2.044	1.27	2.07	1.198	3.07	0.326
2	2.044 [25, 26]	1.27 [25]	2.07 [25]	1.199 [25, 26]	3.07 [25]	0.325 [27]
	1.965 [19]	1.46 [19]	2.01 [19]	1.114 [19]	4.16 [19]	
	4.413	0.324	15.2	2.894	0.577	0.0452
20	4.431 [25, 26]	0.322 [25]	15.1 [25]	2.931 [25, 26]	0.598 [25]	0.0452 [27]
	4.097 [19]	0.507 [19]	15.6 [19]	2.638 [19]	1.05 [19]	

Notes: Our calculations (the top number in each column) are compared with the exact solution of the Schrödinger equation[25–27] and with variational calculations[19] (the literature sources are given in brackets after each entry). To avoid confusion, we make note of the fact that our value of $\mathscr{B}$ is twice the corresponding dimensionless field in the tables of Refs. 25–27.

the corresponding exact results obtained in Refs. 25–27 and with the results of the variational method.[19] We illustrate this comparison for the $1s$ and $2p_{-1}$ states. Table II shows the values of the energy $\mathscr{E}_i$ along with the quantities $\langle i|r^2|i\rangle$ and $\pi|\Psi_{1s}(0)|^2$ and the dipole strength $|\langle 2p_{-1}|\rho e^{-i\varphi}|1s\rangle|^2/2$ for the $1s\rightarrow 2p_{-1}$ transition. Clearly, our results are very close to the exact values, confirming the reliability of the wave functions determined by our method.

The exact results for the energy states in Refs. 28 and 29 almost always agree with the results in Refs. 25–27 in the approximation of interest to us, making it redundant to compare with Refs. 28 and 29.

It is interesting to observe that in the case $\gamma=1$ the binding energy of the $2s$ state and other ''upper'' states of the same symmetry is a nonmonotonic function of the field $\mathscr{B}$ (Refs. 25–29). For example, $\mathscr{B}-\mathscr{E}_{2s}$ has a weak maximum and a weak minimum for $\mathscr{B}\simeq 0.15$ and $\mathscr{B}\simeq 0.3$, respectively. Our calculations confirm this result and also agree with the data of Refs. 25–29 to at least four significant figures. On the other hand, for $\gamma=0.208$ and $\gamma=0.05134$ no such nonmonotonic behavior of the levels of the $2s$ states is observed in the investigated range of dimensionless fields in the two-dimensional problem. A direct calculation shows that when γ deviates from unity, the nonmonotonic behavior of the binding energy for the $2s$ state smooths out and rapidly disappears.

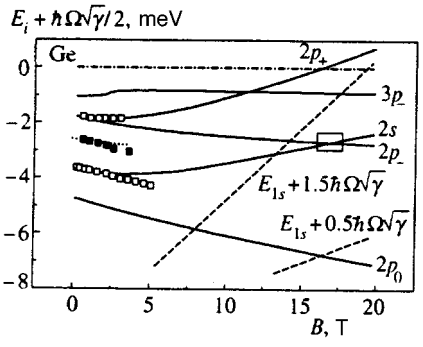

FIG. 3. Energy levels of shallow donors in germanium in a field $\mathbf{B}\|\langle 110\rangle$ for E-type valleys oriented along $\langle -111\rangle$ and $\langle 1-11\rangle$. The energy is read from the edge of the conduction band (zeroth Landau level $\hbar\Omega_E/2=\hbar\Omega\sqrt{\gamma}/2$), which is indicated by a horizontal dot–dash line. The dashed lines represent the energies of the $1s$ state shifted by $\hbar\Omega_E$ and $2\hbar\Omega_E$. Their points of intersection with the solid curves give (in the effective-mass approximation) the points of the principal and second inelastic magnetic–impurity resonance peaks. The point symbols represent experimental data;[13] the graphs are normalized to the $2p_{-1}$ (□) and $2p_0$ (■) levels. The dashed curve represents the conditionally (see text) calculated energy of the $3p_0$ state.

4. LEVELS FOR A VALLEY ORIENTED PERPENDICULAR TO THE MAGNETIC FIELD (THREE-DIMENSIONAL PROBLEM)

The results of calculations of donor states for E-type orientation are shown in Figs. 3 and 4. We note that the corresponding energy levels are situated below those for type-A orientation, so that, in particular, the $1s^{(E)}$ state (see Fig. 5 below) is the principal donor ground state in silicon for $\mathbf{B}\|\langle 001\rangle$ and in germanium for $\mathbf{B}\|\langle 110\rangle$. A distinctive aspect of the three-dimensional problem is the emergence of a characteristic nonmonotonic behavior of the binding energy of the second states in the pairs $(1s,2s)$, $(2p_-,3p_-)$, and $(2p_0,3p_0)$, which in a certain sense has an origin similar to that of the nonmonotonicity mentioned at the end of the preceding section.

For example, let us consider the $2s$ state. For small fields $\mathscr{B}$ the binding energy obviously exhibits a positive shift, linear in the field, due to a displacement of the zeroth Landau level, along with a ''diamagnetic'' quadratic shift. The fact that the quadratic shift is toward lower binding energy is

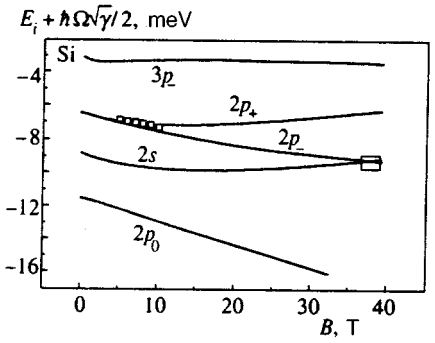

FIG. 4. Energy levels of shallow donors in silicon, read from the edge of the conduction band, in a field $\mathbf{B}\|\langle 001\rangle$ for four E-type valleys oriented along $\langle\pm 100\rangle$ and $\langle 0\pm 10\rangle$. The point symbols represent experimental data;[11,12] the graphs are normalized to the $2p_-$ level.

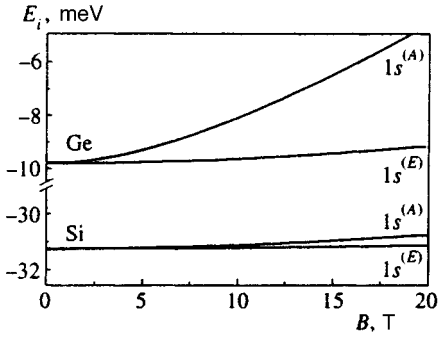

FIG. 5. Energy levels of the ground state calculated in the effective-mass approximation for A- and E-valleys in germanium and silicon. In every case the energies are read from the edge of the conduction band of the corresponding semiconductor in zero magnetic field.

dictated by a general consideration: The introduction of the field **B** in the system lowers the symmetry of the Hamiltonian and the Schrödinger equation (2.13). In the preceding section we have seen that if the value of γ differs markedly from unity, the nonmonotonic behavior disappears. In the two-dimensional problem, in fact, the field **B** does not alter the symmetry of the Schrödinger equation, and the quadratic shift for small $\mathscr{B}$ is not too great. The problem comes closer to the two-dimensional version for higher values of $\mathscr{B}$, because the role of the ''deformed'' Coulomb potential in the Schrödinger equation (2.13) becomes smaller than the term $\mathscr{B}'^2(x^2+y'^2)/4$. As $\mathscr{B}$ increases, the solution should approach the solution for the hydrogen atom in the renormalized magnetic field $\mathscr{B}'$. Now, of course, the binding energy slowly (logarithmically) increases. All the same, the minimum of the binding energy for the $2s$ state is attained in germanium and in silicon for physically unattainable fields and falls outside the limits of Figs. 3 and 4. For the $3p_-$ state, on the other hand, both extrema fall within a physically significant range.

A special approach to the three-dimensional problem is required in calculating the wave function and the energy of the $2p_+$ state [see Eqs. (2.14) and (2.15). This state is the ''upper'' state in its pairing with $2p_0$. Moreover, the same symmetry is possessed by a state that is naturally labeled with the index $3p_0$ and must be constructed by means of a trial wave function having a preexponential factor of the form $(h'x-iy')Q_{3p_0}$ with the factor Q (2.16). It goes over to the two-dimensional $3p_0$ state for $\mathscr{B}=0$ $(h'=0)$ or to the two-dimensional $3p_{-1}$ state for $\gamma=1$ (in which case $h'=1$). In the latter case the $2p_{+1}$ and $3p_{-1}$ levels cross at $\mathscr{B}\simeq 0.13$ (Refs. 27 and 29). As γ deviates from unity, the crossing changes to an anticrossing, and the states go over to upper hybrid and lower hybrid states, respectively. The anticrossing zone, on the one hand, tends to shift to the right with respect to $\mathscr{B}$, since the cyclotron energy decreases (in dimensionless units it is $2\mathscr{B}'=2\mathscr{B}\sqrt{\gamma}$), and the $2p_+$ level therefore becomes lower. But then a decrease in γ causes the energy $\mathscr{E}_{3p_0}$ for small $\mathscr{B}$ (to the left of the anticrossing zone) to merge with $\mathscr{E}_{2p_+}$ and then to drop below the $2p_+$ level and to do so over the entire range of fields $\mathscr{B}$. This inversion effect occurs in germanium (see Fig. 1). For silicon, on the other hand, our calculations show that the anticrossing zone shifts toward higher fields (approximately in the vicinity of $\mathscr{B}\simeq 1.2$), so that in the most relevant range for us the lower hybrid state is in fact the $2p_+$ ground state. A decrease in γ then causes the anticrossing zone to shift rapidly to the left until it reaches zero, now the crossing is no longer present in Ge, and the hybridization of the $2p_+$ and $3p_0$ states must be small.

If at the outset we consider the triple of states $(2p_0,2p_+,3p_0)$ rather than a pair, the Ritz method can be used, in principle, to find both hybrid states. Invoking the method described Sec. 2 in its literal form, we determine only the lower hybrid state, so that this method needs to be corrected for calculating the $2p_+$ state in Ge. To ''jump'' through the $3p_0$ state without calculating it and to go directly to the calculation of the $2p_+$ state in the two-dimensional basis of the functions (2.14)–(2.16), we refrain from varying the parameter g and merely assume that $g=h$. The parameter h is evaluated in the first stage of solving the variational problem, the calculations confirming that $h\to 0$ as $\mathscr{B}\to 0$ and that $h\to 1$ as $\mathscr{B}\to\infty$. As a result, the $2p_+$ state also automatically has the correct asymptotic forms of the wave functions. This is the technique used to obtain our results describing the given state in germanium.

For silicon, in keeping with the foregoing discussion, we can proceed by analogy with Sec. 2. The resulting $2p_+$ state in Si is essentially the lower hybrid state far from the anticrossing zone. Applying the same computational procedure verbatim to the case $\gamma=0.05134$, we obtain an energy close to the correct value of $\mathscr{E}_{3p_0}$ in Ge for small values of $\mathscr{B}$. However, the error of determination of the wave function without regard for the mandatory classification of the solution with respect to principal quantum number (2.8) is not very acceptable in this case. The patently incorrect asymptotic behavior of the preexponential factor when squared terms are ignored produces large errors in the calculation of the matrix element and also in the calculation of the energy for large $\mathscr{B}$. Nonetheless, the dotted line in Fig. 3 is shown to represent the $3p_0$ level found by this approach; as $\mathscr{B}\to 0$, our result goes over to the corresponding result for the two-dimensional problem in the preceding section with three-place accuracy.

Also indicated in Fig. 3 are the experimental values of the energy in the $2s$ and $2p_+$ states, obtained from measurements of the Zeeman components of the $2s\to 2p_-$ and $2s\to 2p_+$ transitions (Ref. 13) with normalization to the energy of the $2p_-$ state. The appreciable discrepancy with experiment for the $2s$ state in Ge can be attributed essentially to the same causes as those discussed in the preceding section. The dark squares in this figure represent experimental data on the position of the $3p_0$ level, determined by subtracting the energies of the $2p_0\to 3d_{+1}$ and $3p_0\to 3d_{+1}$ transitions (see Ref. 13) with normalization, naturally, to the $2p_0$ level. It is interesting to notice that in Ref. 10 splitting of the $2p_\pm$ line in Si for the E-orientation was not observed at all up to fields of 6 T. Our calculations show that such splitting should be appreciable under present-day experimental conditions, beginning with fields as low as 4 T, consistent with the measurements reported in Refs. 11 and 12. The energy val-

ues in the $2p_+$ state for $B=5-10$ T, which can be obtained from graphs in Refs. 11 and 12 with normalization to the $2p_-$ level, are indicated in Fig. 4.

We shall not give detailed results of calculations of the energy in the $1s$ ground state, because the corresponding values determined in the effective-mass approximation, as mentioned, are too remote from reality. However, within this approximation it is interesting to trace the extent to which the energy E_{1s} depends on the orientation of the field. A comparison of the energies in the $1s^{(A)}$ and $1s^{(E)}$ states is illustrated in Fig. 5. As in the preceding section, small boxes enclose the points of elastic magnetic-impurity resonance in Figs. 3 and 4.

5. SCATTERING OF A CONDUCTION ELECTRON BY AN EXCITED NEUTRAL DONOR, AND ELASTIC MAGNETIC-IMPURITY RESONANCE

In principle, the scattering problem can be analyzed for both types of carriers, but the lifetime of free holes in n-type semiconductors is too short. The main effect of injecting holes into our system would be lowering of the Fermi level and donor ionization, which naturally preclude the possibility of observing the resonance.

We also note that, being concerned with temperatures $T\lesssim 10$ K and fields $B\gtrsim 10$ T, we interpret electrons of the zeroth Landau band, specifically those associated with valleys having the lowest of all cyclotron frequencies, as free carriers. For silicon with $\mathbf{B}\|\langle 001\rangle$ these particles are $\langle \pm 100\rangle$ and $\langle 0\pm 10\rangle$ quadruplet electrons, for germanium with $\mathbf{B}\|\langle 111\rangle$ they are triplet-forming valleys, and for germanium with $\mathbf{B}\|\langle 110\rangle$ they are associated with the lowest doublet (see Appendix 1). As we shall see below, the characteristic length governing the main scattering contribution is the magnetic length $\lambda=\sqrt{\hbar c/eB}$, so that the problem of scattering by an impurity is solvable in the long-range approximation if λ is much greater than the characteristic Bohr radius in the material:

$$\lambda \gg a_B. \tag{5.1}$$

For silicon near resonance values of the field this condition is satisfied with a fair margin [see (2.2)], whereas for germanium λ and a_B are obviously of the same order. This fact does not imply that the effect *per se* does not occur in germanium, but one must bear in mind that the result obtained below in application to Ge is valid only in order of magnitude.

The degeneracy at the crossing point of the $2s$ and $2p_-$ levels is lifted if the perturbing potential is taken into account. If the impurity is excited and resides in one of the two states resulting from $2s-2p_-$ hybridization, the energy of a bound electron depends (linearly in the vicinity of the crossing point) on the electric field generated by a moving free carrier. On the other hand, the combined energy of the neutral donor and the carrier is determined entirely by their relative spatial position, and the conduction electron can therefore be regarded as acted upon by the scattering potential of the donor field. By this reasoning, adiabaticity of the action of the conduction electron on the bound electron admits an obvious interpretation, so that the eigenstate problem for each of them can be solved in succession utilizing the ''slowness'' of the variables (both spatial and temporal) describing the state of the conduction electron. An elementary estimate shows that adiabaticity holds in the vicinity of the point of degeneracy if the free-carrier energy ε relative to the zeroth Landau level satisfies the inequality

$$\epsilon \ll \frac{m_z e^4}{\hbar^2\kappa^2}\left(\frac{a_B}{r}\right)^2 \sim \frac{m_z e^4}{\hbar^2\kappa^2}\left(\frac{a_B}{\lambda}\right)^2 = \hbar\Omega_z, \tag{5.2}$$

where m_z is the effective mass of longitudinal motion along $\mathbf{B}$, Ω_z is the corresponding cyclotron frequency, and $\mathbf{r}$ is the position of the band electron relative to the impurity center. For $\epsilon<10$ K the adiabaticity condition can be regarded as fulfilled.

In the remainder of this section, unless otherwise stated, we consider quantities to have their customary units. As before, we place the origin at the site of the central unit cell containing the impurity ion. Let $\tilde{\mathbf{r}}$ and $\mathbf{r}$ denote the coordinates of the bound electron and the conduction electron, respectively; the Coulomb energy of interaction of the donor with the conduction electron is then given by the following relation in the dipole approximation:

$$\frac{e^2}{\kappa}\left(\frac{1}{|\mathbf{r}-\tilde{\mathbf{r}}|}-\frac{1}{r}\right)\approx\frac{e^2\tilde{\mathbf{r}}\cdot\mathbf{r}}{\kappa r^3}, \quad r\gg\tilde{r}. \tag{5.3}$$

From this result, using a standard perturbation procedure, we find the energy in the vicinity of the degeneracy point as a function of the adiabatic variable $\mathbf{r}$:

$$E_\pm = \pm\sqrt{\frac{(\Delta E)^2}{4}+\frac{e^4|\mathbf{d}\cdot\mathbf{r}|^2}{\kappa^2 r^6}}+\frac{E_{2s}+E_{2p_-}}{2}, \tag{5.4}$$

where $\Delta E(B)=E_{2s}(B)-E_{2p_-}(B)$ is the energy difference between unperturbed levels, and $\mathbf{d}=\langle 2s|\mathbf{r}|2p_-\rangle$ is a matrix element, in which the angle brackets signify convolution with respect to the coordinates of the bound electron. From the energy (5.4) it is now required to subtract the donor energy at $|\mathbf{r}|=\infty$, i.e., the quantity E_{1s} or E_{2p_-}, depending on the sign in front of the square root. We then obtain the effective conduction electron–donor interaction potential

$$U_\pm(\mathbf{r})=\pm\left(\sqrt{\frac{(\Delta E)^2}{4}+\frac{e^4|\mathbf{d}\cdot\mathbf{r}|^2}{\kappa^2 r^6}}-\frac{|\Delta E|}{2}\right). \tag{5.5}$$

We recall that the quantities involved in this expression depend parametrically on the magnetic field and, in addition, that ΔE vanishes at the degeneracy point. We denote the corresponding field by B_0 (or by $\mathscr{B}_0$ in dimensionless form). We have already seen that the field B_0, in turn, depends on the orientation of the conduction-electron valley relative to the vector $\mathbf{B}$. Equations (5.4) and (5.5) are valid if the distance from other levels is sufficiently large:

$$|\Delta E| \ll |E_{2s}-E_i|, \quad i\neq 2p_-, 2s \tag{5.6}$$

(the strength of the dipole transition from the $2s$ or $2p_-$ state to other states is assumed to be lower than or of the same order as $|\mathbf{d}|^2$). The sign of the interaction potential (5.5) de-

pends on the particular hybridized state, lower or upper, occupied by the donor; the sign of U does not affect our subsequent results.

We now proceed to calculate the characteristics of the scattering of an electron by the impurity potential (5.5). Irrespective of the orientation of the effective-mass ellipsoid, we choose the z axis parallel to **B**, and align the y axis with the field **F**. In the first nonvanishing scattering approximation the quantum-mechanical result for the transverse diagonal conductivity is known[38] to coincide with the quasiclassical expression (see Refs. 36 and 37)

$$\sigma_{yy}=-\frac{e^2}{2V}\sum_{\nu\nu'}\frac{df}{d\epsilon}(y_0-y_0')^2W_{\nu\nu'}, \tag{5.7}$$

which is easily obtained by analyzing the transverse diffusion of the Larmor ''disks'' (Ref. 35). In this equation the indices ν and ν' denote the complete sets of quantum numbers characterizing the initial and final (after scattering) states of the electron, $W_{\nu\nu'}$ is the scattering probability, $f(\epsilon)$ is the energy distribution of electrons in the conduction band (in equilibrium f is the Fermi function), and $V=L_xL_yL_z$ is the volume of the sample. Equation (5.5) assumes that the representation of the electron states is chosen in a form such as to preserve y_0 and y_0', i.e., so they can be expressed in terms of the corresponding set ν or ν'. In the classical case y_0 is the y coordinate of the Larmor disk, and in the quantum case it is the equilibrium position of a magnetic Landau oscillator. Consequently, the representation of the conduction-electron wave function is found by means of the Landau gauge ($A_x=-B_y$, $A_y=A_z=0$), where it is of utmost importance that the choice of the y axis is strictly regulated by the direction of the field **F**. In the isotropic case, of course, this consideration is irrelevant, but in the problem with an anisotropic effective mass we must allow for the fact that the wave free-electron wave function in a magnetic field now depends on the direction of **F**, even if the magnitude of the electric field itself is negligible in this case.

The solution of the Schrödinger equation also reduces to oscillator functions for arbitrary orientation of the axes of the effective-mass ellipsoid relative to **B** and **F**. Only the magnetic length is renormalized as a function of this orientation. However, we still simplify the problem. First of all, we assume that the electric field is weak:

$$eF\lambda\ll\hbar\Omega_z; \tag{5.8}$$

this condition is usually satisfied when conduction electrons are present in the material (i.e., for characteristic fields $F\lesssim 10$ V/cm) and implies that F can be neglected in the Schrödinger equation.[5)] Second, we confine our discussion to the case in which the field **B** is directed along one of the principal axes of the effective-mass tensor, i.e., when $m_z=m_\perp$ and $m_{xz}^{-1}=m_{yz}^{-1}=0$. Finally, recognizing that the relation $\epsilon\ll\hbar\Omega_z$ holds by virtue of (5.2), we retain only the one zeroth Landau level. As a result, the set ν is determined entirely by the wave vectors k_x and k_z, and the energy $\hbar\Omega_E/2+\epsilon$ of the free electron is determined entirely by the quantum number k_z (here $\epsilon=\hbar^2k_z^2/2m_\perp$, and $\Omega_E=\Omega\sqrt{\gamma}$ is the cyclotron frequency, which in the given situation corresponds to TM orientation). The band-electron wave function

$$\Psi_\nu=\frac{[\mathrm{Re}\ \eta(\alpha)/\pi]^{1/4}}{\sqrt{L_xL_z\lambda}}\exp\left[-\frac{\eta(\alpha)(y-y_0)^2}{2\lambda^2}+ik_xx+ik_zz\right],\quad y_0=\hbar k_xc/eB, \tag{5.9}$$

depends on the angle α formed by the y axis with the symmetry axis of the effective-mass ellipsoid. In Eq. (5.9) this dependence is specified by the complex function

$$\eta(\alpha)=\frac{\sqrt{\gamma}-i(1-\gamma)\sin 2\alpha/2}{\sin^2\alpha+\gamma\cos^2\alpha}. \tag{5.10}$$

The scattering probability is given by the well-known expression

$$W_{\nu\nu'}=\frac{2\pi}{\hbar}|U_{\nu\nu'}|^2\delta(\epsilon-\epsilon'), \tag{5.11}$$

where

$$U_{\nu\nu'}=\left\langle\nu\left|\sum_i U(\mathbf{r}-\mathbf{R}_i)\right|\nu'\right\rangle \tag{5.12}$$

is the matrix element of the total perturbation of impurities situated at the points $\mathbf{R}_i$, and $U(\mathbf{r})$ is given by Eq. (5.5). If the impurities are distributed randomly in space, the substitution of Eq. (5.12) into (5.11) and summation on i verify that the ''interference'' terms are vanishingly small, so that the operations of summation and squaring can be interchanged. Moreover, the sum over the impurities must obviously be replaced by integration according to the rule $\Sigma_i\to\int n_d^*d\mathbf{R}=n_d^*L_xL_z\int dY$, where n_d^* is the density of excited donors, which is assumed to be homogeneous. As a result, substituting Eq. (5.9) into (5.12)), we obtain

$$W_{\nu\nu'}=\frac{2\hbar^3n_d^*\lambda}{m_\perp^2L_xL_z}\delta(\epsilon-\epsilon')\mathscr{S}\left(\frac{y_0-y_0'}{\lambda},\lambda k_z-\lambda k_z'\right), \tag{5.13}$$

where the dimensionless function

$$\mathscr{S}(v,q)=\varrho\ \exp\left(-\frac{|\eta|^2v^2}{2\varrho}\right)\int_{-\infty}^{\infty}du\left|\int d^3\overline{\mathbf{r}}\,\mathscr{U}(\overline{\mathbf{r}})\times\exp\left[-iv\overline{x}-iq\overline{z}-\varrho\left(\overline{y}-u+\frac{iv\vartheta}{2}\right)^2\right]\right|^2 \tag{5.14}$$

has been constructed by transforming to the dimensionless variables $\mathbf{r}-\mathbf{R}=\lambda\overline{\mathbf{r}}$, $y_0-Y=\lambda(u+v/2)$, $y_0'-Y=\lambda(u-v/2)$, $k_z-k_z'=q\lambda^{-1}$ and to the dimensionless potential $\mathscr{U}=U/\hbar\Omega$ and then replacing the integration over Y by integration over u. In Eq. (5.14) we have also introduced the notation $\varrho(\alpha)=\mathrm{Re}\ \eta(\alpha)$ and $\vartheta(\alpha)=\mathrm{Im}\ \eta(\alpha)/\mathrm{Re}\,\eta(\alpha)$.

It is important to note that here we have already incorporated the Born approximation in the calculations; in other words, in Eq. (5.12) both the ''initial'' state $|\nu\rangle$ and the ''final'' state $\langle\nu'|$ are assumed to be unperturbed, i.e., specified by the functions (5.9). The admissibility of this approximation must be justified; in this regard, see Appendix 3.

5.1. Lifetime of the band state (longitudinal scattering characteristic)

Elasticity in reference to scattering implies that the quantity q is equal to zero or to $2\lambda k_z$. It is also obvious that a nonzero value of q always introduces a smaller contribution to the reciprocal lifetime of the state ν of the band electron,

$$\frac{1}{\tau}=\sum_{\nu'}^{\nu'\neq\nu} W_{\nu,\nu'}, \tag{5.15}$$

than the contribution of scattering with k_z conserved. On the other hand, it follows from condition (5.2) that the possible values of q are small ($q\ll 1$) in any case, so that scattering with the sign of k_z unchanged and with a change of sign to leading order are equiprobable. We shall assume everywhere below that $q=0$.

Before calculating the transverse conductivity, we determine the quantity (5.15). We do so initially in the immediate vicinity of the degeneracy point B_0, assuming that

$$\mathscr{U}\approx\frac{e^2|\mathbf{d}\cdot\overline{\mathbf{r}}|m_\perp}{\kappa\hbar^2\overline{r^3}}. \tag{5.16}$$

As a result,

$$\int \mathscr{U}(\overline{\mathbf{r}})d\overline{z}=\frac{\mathscr{D}}{\sqrt{\overline{x^2}+\overline{y^2}}}, \tag{5.17}$$

where $\mathscr{D}(\mathbf{B})$ is the dimensionless matrix element

$$\mathscr{D}=\begin{cases}\langle 2s|\rho e^{i\varphi}|2p_{-1}\rangle & \text{for } A\text{-ellipsoid (see Sec. 3)},\\ 2\langle 2s|z|2p_{-1}\rangle & \text{for } E\text{-ellipsoid (see Sec. 4)}.\end{cases} \tag{5.18}$$

The angle brackets are employed in (5.17) with the same sense as, for example, in Eq. (3.1), i.e., they signify averaging in the dimensionless basis x,y,z used in the preceding sections. Consequently, the calculation of $\mathscr{D}$ simply represents the case in which it is necessary to know the wave functions of the bound electron states of the donor. Our calculations show that the dimensionless quantity (5.18) has the following values exactly at the point $B=B_0$:

for Ge: $|\mathscr{D}_A|^2=0.25$ for $\mathscr{B}_0^{(A)}=1.50$,

$|\mathscr{D}_E|^2=8.2$ for $\mathscr{B}_0^{(E)}=2.54$;

for Si: $|\mathscr{D}_A|^2=3.73$ for $\mathscr{B}_0^{(A)}=0.16$,

$|\mathscr{D}_E|^2=16.8$ for $\mathscr{B}_0^{(E)}=0.575$.

We note that in making the transition to (5.17) we have disregarded the quantity ΔE in Eq. (5.5) in the derivation. This means that the integration over $\overline{\mathbf{r}}$ in (5.14) makes sense if the main contribution is from values

$$\overline{r}\ll\sqrt{|\mathscr{D}|\hbar\Omega/|\delta E|}. \tag{5.19}$$

Here $\delta E=\max(\Delta E, eF\lambda)$. A different estimate of the upper bound of $\overline{r}$ follows from the adiabaticity condition (5.2) and in fact, since $\mathscr{D}\sim 1$, simply entails the substitution of ϵ for δE in (5.19). Naturally, $\overline{r}$ is also bounded above by the interimpurity distance:

$$\overline{r}\ll n_d^{*-1/3}/\lambda. \tag{5.19'}$$

To calculate the reciprocal scattering time (5.15), we change from summation over ν' to integration: $\Sigma_{\nu'}=\int d\epsilon' g(\epsilon')\int dy_0'$, where

$$g(\epsilon)=\frac{4L_xL_z\sqrt{m_\perp/2\epsilon}}{(2\pi\lambda)^2\hbar} \tag{5.20}$$

is the density of quantum states in the zeroth Landau band per unit energy and unit length along y. The factor 4 in Eq. (5.20) corresponds to the two possible signs of k_z for a given value of ε and two spin states. As a result,

$$\frac{1}{\tau(\epsilon)}=\frac{\sqrt{2}\hbar^2 n_d^*}{\pi^2 m_\perp^{3/2}\sqrt{\epsilon}}\int_{-\infty}^{\infty} dv\,\mathscr{S}(v,0), \tag{5.21}$$

and in the expression for $\mathscr{S}(v,0)$ it is required to use the result of integration over $\overline{z}$ (5.17), (5.18). We then find that formally the integral (5.21) diverges logarithmically for small values of $|v|$, since the equation (A2.8) obtained in Appendix 2 is valid for it in this case. Of course, the divergence is easily removed: The integration over $\overline{x}$ to obtain Eqs. (A2.3) and (A2.4) rests on the assumption that the condition $|v\overline{x}|\gg 1$ holds at the boundaries of the domain of integration, so that the cutoff at small $|v|$ occurs at values of $\overline{v}$ reciprocal to the quantity on the right-hand side of inequalities (5.19) and (5.19'). As a result, the time between collisions can be determined with logarithmic accuracy:

$$\frac{1}{\tau(\epsilon)}=\frac{2^{5/2}\pi|\mathscr{D}(\mathbf{B}_0)|^2\hbar^2 n_d^*}{\sqrt{m_\perp^3\epsilon}}\Lambda, \tag{5.22}$$

where $\Lambda=-\ln(\lambda n_d^{*1/3}+\delta E/|\mathscr{D}|\hbar\Omega)$, and $\delta E=\max\{|E(B)-E(B_0)|, |eF\lambda|\epsilon\}$.

Curiously, the collision frequency (5.22) does not contain the interaction constant. This quantity (e^2/κ) ''cancels out,'' because it occurs twice in expressions (5.4), (5.5), and (5.16): once as a factor governing the perturbation field of the band electron and again as a divisor attributable to the fact that $|\mathbf{d}|\sim a_B\propto\kappa/e^2$ holds for the bound electron. The interaction constant itself does not enter into the answer only so long as the long-range condition (5.1) is satisfied.

A pivotal consideration is the fact that for the final results to be valid, the time τ must have both upper and lower bounds. The upper bound is associated with the assumption that the given scattering channel is actually the dominant one. Other channels inevitably present in our system are scattering by acoustic phonons and scattering by neutral but unexcited donors, whose number n_d can be higher than n_d^*, but the potential of the perturbation created by them has a Gaussian shape. Note that the lifetime of the band state is of the same order as the relaxation time of the longitudinal momentum (see Ref. 2, Sec. 14.1). We know[2,37,38] that the latter time is weakly dependent on the magnetic field and can be estimated from the experimental values of the mobility in a nondegenerate semiconductor for $\mathbf{B}=0$. At a temperature $T<10$ K the mobility attains values of 10^5 cm$^2\mathbf{B}^{-1}$s^{-1} in silicon if $n_d\lesssim 10^{14}$ cm^{-3}, and it attains 10^6 cm$^2\mathbf{B}^{-1}$s^{-1} in germanium for $n_d<10^{13}$ cm^{-3} (see Refs. 48 and 49 and the literature cited therein). Consequently, the characteristic re-

laxation time of the longitudinal momentum τ_0 as determined by other scattering channels is on the order of 10 ps, which is fully consistent with the theoretical calculations of τ_0 in the ultraquantum case of *DA* scattering by phonons (Ref. 2, Sec. 14.3), i.e., for $\epsilon \gtrsim \hbar s/\lambda$ (s is the speed of sound). The inequality $\tau < \tau_0$ imposes a condition on the density of excited impurities; we infer from (5.22) that the relation $n_d^* \gtrsim 10^{13}$ cm^{-3} must hold for Si, and the relation $n_d^* \gtrsim 10^{11}$ cm^{-3} for Ge. In this sense germanium is the preferred material under conditions such that $n_d^* < n_d$.

In addition to the foregoing estimate, which is based on experiment, it is useful to compare Eq. (5.22) with the theoretical relaxation time determined by scattering at a neutral, unexcited donor. Unfortunately, since the long-range approximation is inapplicable, this problem is very complicated and has therefore received little attention to date. All the same, for estimates we can use results obtained on the basis of Erginsoy's model.[39] The present topic of concern, of course, is the non-Born case and treats the problem of the scattering of a hydrogenlike atom (with the dielectric permittivity and the effective mass introduced into the model) with allowance for exchange interaction between the scattered and bound electrons. The particle interaction zone is of order a_B. For our estimations, however, it is still more convenient to define a certain effective scattering potential, which in the Born case for zero magnetic field gives the same result for the cross section as in Ref. 39, whereupon the scattering problem in a quantizing [see (5.2)] magnetic field can be solved for this potential. This is the scheme used in Ref. 50 to calculate the relaxation time for a neutral impurity. Accordingly, as in the case of (5.22), this time (which we denote by τ_{Er}) is proportional to $\sqrt{\epsilon}$,[6)] and the ratio of the two times is

$$\frac{\tau(\epsilon)}{\tau_{Er}(\epsilon)} = \frac{15 n_d a_B}{64\sqrt{2\pi}|\mathscr{D}|^2 \Lambda n_d^* \lambda} \tag{5.23}$$

(here we disregard the anisotropy of the effective mass). In the cases of interest to us, therefore, Erginsoy relaxation is at least two orders of magnitude slower, and the long-range effect is already appreciable beginning with $n_d^*/n_d \sim 0.01$.

Finally, we determine the lower bound on τ. It is associated with the applicability of the time-dependent perturbation equation (5.11) and stems from the requirement $\tau > \delta\tau$, where $\delta\tau$ is the quantum-mechanical time uncertainty, i.e., $\delta\tau \sim \hbar/\overline{\epsilon} > \hbar/\epsilon$. The quantity $\overline{\epsilon}$ represents the lowest energy at which our theory is valid. Combining the given inequalities and taking (5.22) into account, we find that

$$\epsilon > \overline{\epsilon} \sim \frac{\hbar^2 (n_d^* \Lambda)^{2/3} |\mathscr{D}|^{4/3}}{m_\perp}. \tag{5.24}$$

This result actually implies that the free-electron de Broglie wavelength in the direction of the magnetic field must be bounded above by the mean free path $\tau\sqrt{\epsilon/m_\perp}$.

We can now calculate the reciprocal time (5.21) over a broader neighborhood of resonance without ignoring ΔE in (5.5). Equations (5.16) and (5.17) are then invalid, and the integration in (5.21) cannot be carried out analytically. As a result, we have

$$\frac{1}{\tau(\epsilon)} = \frac{16\sqrt{2}\pi^2\hbar^2 n_d^*}{\sqrt{m_\perp^3 \epsilon}} \mathscr{T}_\parallel(\mathscr{B}). \tag{5.25}$$

The function $\mathscr{T}_\parallel$ is calculated differently, depending on the type, *A* or *E*, of the localized-electron valley:

$$\mathscr{T}_\parallel^{(A,E)}(\mathscr{B}) = \frac{|\mathscr{D}|^2}{2\pi} G_\parallel^{(A,E)}\left(\frac{\Delta\mathscr{E}}{4\mathscr{B}|\mathscr{D}|}\right). \tag{5.26}$$

Here

$$G_\parallel^{(A,E)}(\zeta) = \frac{\rho}{\pi^3} \int_0^\infty dv e^{-\varrho v^2/2} \int_{-\infty}^\infty du \mathscr{L}_{A,E}^2(\varrho, \zeta, u, v), \tag{5.27}$$

$$\mathscr{L}_{A,E} = \int_0^\infty d\overline{x} \cos(v\overline{x}) \int_{-\infty}^\infty d\overline{y} e^{-\varrho(\overline{y}-u)^2} \times \int_0^\infty d\overline{z} \left[\sqrt{\zeta^2 + \frac{s_{A,E}(\overline{\mathbf{r}})}{(\overline{x^2}+\overline{y^2}+\overline{z^2})^3}} - |\zeta|\right], \tag{5.27'}$$

$s_A = \overline{x^2} + \overline{y^2}$, and $s_E = \overline{z^2}$. In Eq. (5.26) $\Delta\mathscr{E}$ is the energy difference ΔE reduced to dimensionless form by the rule (2.2). In the expression (5.14) for $\mathscr{S}(v,0)$ we have set $\alpha = \pi/2$, since it is clear from physical considerations that our result will not depend on the angle α. This can be seen by examination of the integral (5.21). We therefore have $\varrho = \sqrt{\gamma}$ in Eqs. (5.27) and (5.27'). The function (5.27) is also naturally invariant under the substitution $\varrho \to 1/\varrho$.

We note that the time between collisions (5.21), (5.22), (5.25) can be used to find the longitudinal conductivity

$$\overline{\sigma_{zz}} = ne^2 \langle\langle \tau \rangle\rangle / m_\perp, \tag{5.28}$$

where the density of band electrons is

$$n = 4 \int d\epsilon \frac{g(\epsilon)}{L_x L_z} f(\epsilon). \tag{5.29}$$

and the double angle brackets signify energy averaging, which can be carried out here by the same scheme as in zero magnetic field (Ref. 2, Sec. 7.2). The one-dimensionality of the problem in the presence of a quantizing magnetic field yields

$$\langle\langle \tau \rangle\rangle = -\frac{8}{n} \int d\epsilon \frac{df}{d\epsilon} 2\epsilon\tau(\epsilon) \frac{g(\epsilon)}{L_x L_z}. \tag{5.30}$$

The factor 4 in (5.29) corresponds to the four states of the band quadruplet in Si. Allowance must also be made for the fact that in the given approximation (5.2) the frequency of collisions in which the sign of k_z changes is exactly half the total frequency $1/\tau$, thus accounting for the factor 8 in Eq. (5.30).

In our case $f(\epsilon)$ is the distribution of nonequilibrium electrons at the zeroth Landau level. If we assume that it is characterized by an effective temperature, $f = f'(\epsilon/T^*$, then $\overline{\sigma_{zz}} \propto T^*$. The magnetic field dependence is characterized by the function (5.26) and for the case of an *A*-valley is illustrated in Fig. 6.

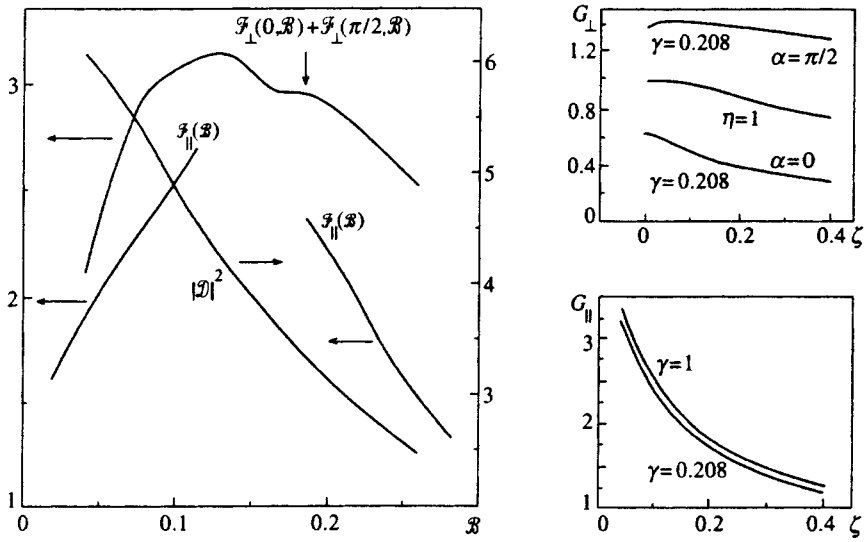

FIG. 6. Results of calculations of dimensionless quantities (see text) governing the behavior of the transport scattering characteristics under elastic magnetic–impurity resonance conditions for the case in which the bound state belongs to an A-valley in silicon (the index A is dropped everywhere). The magnetic field, reduced to dimensionless form according to (2.2), is plotted along the horizontal axis in the main part (left side) of the figure. The point of intersection of the $2s$ and $2p_{-1}$ levels corresponds to $\mathcal{B}=0.16$.

5.2. Calculation of σ_{yy} (transverse scattering)

We substitute Eq. (5.14) into (5.7) and replace the summation by integration:

$$\sum_{\nu\nu'} = \frac{1}{2}\int\int d\epsilon d\epsilon' g(\epsilon)g(\epsilon')\int\int dy_0 dy_0'.$$

The insertion of the factor 1/2 is explained by the fact that the spin state of the conduction electron does not change in scattering, so that the spin doubling of the density of states needs to be taken into account only once in one of the functions $g(\epsilon)$. In Eq. (5.14), as in the calculation of the collision frequency (5.22), it is necessary to set $q=0$. We ultimately arrive at the result

$$\sigma_{yy}(\alpha) = -\frac{e^2\hbar n_d^*}{\pi m_\perp \sqrt{\mathcal{B}}}\mathcal{F}_\perp(\alpha,\mathcal{B})\int_{\epsilon>\overline{\epsilon}}\frac{df}{d\epsilon}\frac{d\epsilon}{\epsilon}. \qquad (5.31)$$

The function of the dimensionless magnetic field $\mathcal{B}$ and the angle α is defined by the expression

$$\mathcal{F}_\perp^{(A,E)}(\alpha,\mathcal{B}) = \sqrt{\mathcal{B}}|\mathcal{D}|^2 G_\perp^{(A,E)}\left(\eta(\alpha),\frac{\Delta\mathcal{E}}{4\mathcal{B}|\mathcal{D}|}\right), \qquad (5.32)$$

where

$$G_\perp^{(A,E)}(\eta,\zeta) = \frac{4\varrho}{\pi^3}\int_0^\infty v^2 dv\,\exp\left(-\frac{|\eta|^2 v^2}{2\varrho}\right) \times\int_{-\infty}^{\infty} du|K_{A,E}(\eta,\zeta,u,v)|^2, \qquad (5.33)$$

$$K_{A,E}(\eta,\zeta,u,v) = \int_0^\infty d\overline{x}\cos(v\overline{x}) \times\int_{-\infty}^{\infty} d\overline{y}\exp\left[-\varrho\left(\overline{y}-u+\frac{iv\vartheta}{2}\right)^2\right] \times\int_0^\infty d\overline{z}\left[\sqrt{\zeta^2+\frac{s_{A,E}(\overline{\mathbf{r}})}{(\overline{x^2}+\overline{y^2}+\overline{z^2})^3}}-|\zeta|\right]. \qquad (5.33')$$

Of course, the transverse conductivity depends on the orientation of the band ellipsoid, i.e., on the angle α.

If $\Delta\mathcal{E}=0$, the expression for the function $G(\eta,0)$ is simplified considerably and does not depend on the type of valley, A or E [see Eqs. (A2.9) and (A2.10) in Appendix 2]. Also, in principle, the situation arises where one can assume that $\eta=1$ [in which case $G(1,0)\equiv 1$; see (A2.11)]. This is the case if the conduction electron belongs to the zeroth Landau level of an A-valley but not to the E-valley discussed so far. Such band states are higher in energy by the amount $\hbar\Omega(1-\sqrt{\gamma})/2$, but are metastable (the ''dumping'' of an electron to the bottom of the conduction band is determined by intervalley transition).

The integrals (5.33)–(5.33′) converge, so that the magnetic length at which cutoff takes place determines the characteristic spatial domain around the impurity center responsible for the main contribution to the transverse conductivity. As mentioned already, at distances $\lesssim a_B$ Eq. (5.5) is invalid, and long-range influence is nonexistent. Naturally we cannot properly take into account the scattering in this central part. If we assume that the energy of interaction of carrier and donor in the given domain remains equal in order of magnitude to the ''atomic energy'' $e^2/\kappa a_B$ (and there is no physical justification for the reverse proposition), we readily conclude from an analysis of (5.33) and (5.33′) that the contribution to σ_{yy} from the central (and clearly non-Born) part of the scattering is small in comparison with (5.31), of order of the dimensionless parameter a_B/λ. We note that in

calculating the time (5.22) the corresponding ratio of the contributions is even smaller, being equal to $a_B/\lambda\Lambda$. Hence, longitudinal scattering is, so to speak, an even longer-range phenomenon.

It is interesting that the contribution of distances $\lesssim a_B$ can also be estimated independently on by way of the following considerations. Let us assume that the long-range dipole component of the potential is absent. Such will be the case, for example, if the state of the donor is the pure $2p_-$ or $2s$ state. The calculation of the scattering characteristics in this case determines the contribution of the central part. For estimates we can once again proceed from the Erginsoy model, but with the one difference that the initial (prescattering) state of the impurity is the corresponding excited state rather than the ground state of the ''hydrogen atom.'' As a result, for the ratio of the lifetimes we obtain an equation of the same order of magnitude as (5.23), in which it is required only to set $n_d^*/n_d=1$. For the transverse conductivity an estimate based on Eqs. (5.14) and (5.7), in which the role of the potential $\mathscr{U}$ must be filled by an appropriate effective potential,[50] shows that the above estimate of the contribution a_B/λ is also confirmed.

We have evaluated the function $G_\perp^{(A)}(\alpha,\zeta)$ for the two directions corresponding to $\alpha=0$ and $\alpha=\pi/2$. In both cases we have $\vartheta=0$, but $\eta(0)=\varrho(0)=1/\sqrt{\gamma}$ and $\eta(\pi/2)=\varrho(\pi/2)=\sqrt{\gamma}$. The results of the calculations are shown in the upper right part of Fig. 6. We have assumed $\gamma=0.208$, i.e., we have treated the case of silicon in a field $\mathbf{B}\|\langle 001\rangle$ (see Fig. 2). Also shown for comparison are the results of calculations for $\eta=1$. The lower right part of Fig. 6 shows a plot of $G_\parallel^{(A)}(\zeta)$ for $\varrho=\sqrt{0.208},1$. It is evident here that the dependence on the parameter γ is even less significant and, of course, this quantity is completely independent of the angle α.

Note that Eq. (5.31) describes the conductivity contribution of only one of the valleys. If all valleys of the band quadruplet are taken into account, we obtain the total conductivity

$$\overline{\sigma_{yy}}=2\sigma_{yy}(\alpha)+2\sigma_{yy}(\pi/2-\alpha). \tag{5.34}$$

The energy integral (5.31) is readily computed in the logarithmic approximation by introducing an effective temperature T^* characterizing the distribution of nonequilibrium band electrons: $(df/d\epsilon)|_{\epsilon=0}=-f(0)/T^*$. Finally, for Si with $\mathbf{B}\|\langle 001\rangle$ we find

$$\overline{\sigma_{yy}}(\alpha)=\frac{2e^2\hbar n_d^* f(0)}{\pi m_\perp T^*\sqrt{\mathscr{B}}}\ln\left(\frac{T^*}{\overline{\epsilon}}\right)\left[\mathscr{F}_\perp(\alpha,\mathscr{B})+\mathscr{F}_\perp\times\left(\frac{\pi}{2}-\alpha,\mathscr{B}\right)\right]\quad\text{for } T^*\gg\overline{\epsilon}. \tag{5.35}$$

The angle $\alpha=0$ corresponds here to the direction $\mathbf{F}\|\langle 010\rangle$ or $\mathbf{F}\|\langle 100\rangle$. As a result, the dependence of $\overline{\sigma_{yy}}$ on the angle α is weak [in a special case this fact is confirmed by Eq. (A2.10) for $G_\perp(\eta,0)$]. In the main part of Fig. 6 for the case of an A-valley we see graphs of $\mathscr{F}_\parallel(\mathscr{B})$ and $\mathscr{F}_\perp(0,\mathscr{B})+\mathscr{F}_\perp(\pi/2,\mathscr{B})$ along with a graph of $|\mathscr{D}(\mathscr{B})|^2$ calculated according to the rule (5.18).

The ratio of the transverse to the longitudinal magnetoresistance can be determined from Eqs. (5.22), (5.25), (5.28), and (5.35): $\rho_{yy}/\rho_{zz}=\sigma_{yy}\sigma_{zz}/\sigma_{xy}^2\sim(e^2/\lambda\Lambda\kappa T^*)\ln(T^*/\overline{\epsilon})$.

If in a semiconductor with unexcited donors the transverse conductivity is equal to $\overline{\sigma_{yy}^{(0)}}$ and the relaxation time of the longitudinal momentum is τ_0, as a result of excitation we obtain $\overline{\sigma_{yy}}/\overline{\sigma_{yy}^{(0)}}\sim\tau_0/\tau$, i.e., the gain is determined by the first power of the ratio of the times between collisions and may not be too large in reality. Far more important is the occurrence of magnetic-impurity resonance in the unusual functional dependence of $\overline{\sigma_{yy}}$ on B. Let the Hall conductivity in our material be $\overline{\sigma_{xy}}$; the transverse magnetoresistance is then

$$\rho_{yy}(B)=\overline{\sigma_{yy}}/\overline{\sigma_{xy}^2}\propto\mathscr{B}^{3/2}[\mathscr{F}_\perp(\alpha,\mathscr{B})+\mathscr{F}_\perp(\pi/2-\mathscr{B})].$$

Ignoring the nonmonotonic behavior described by the function $\mathscr{F}_\perp(\alpha,\mathscr{B})$, we see that, roughly speaking, the magnetoresistance increases with the field as $B^{3/2}$. Our case is in some measure intermediate between scattering by charged impurities (where ρ_{yy} is independent of B; see Ref. 2, Sec. 14.3) and scattering by point defects ($\rho_{yy}\propto B^2$; see the same location).

The main distinctive feature, of course, is the nonmonotonic field dependence of the quantities ρ_{zz} and $\rho_{yy}/B^{3/2}$, with a maximum in the vicinity of $B_0=10.5$ T.

Note that the quantity $|\mathscr{D}|$ increases considerably upon transition from the A-orientation to the E-orientation. This jump is particularly abrupt and conspicuous for germanium. In addition, it is evident from Fig. 3 that the dependence of ΔE on $B-B_0$ is also ''steeper'' than for the A-orientation. The sum total of these considerations leaves us no choice but to assume that in experimental work the investigated resonant effect will (if the magnetic field is directed along $\langle 110\rangle$) be very strong in this semiconductor, despite the reality that condition (5.1) does not hold. The states of both the free and the localized electron in this case are associated with valleys of the E type and are described by Wannier functions of the lower valley-orbital doublet (see Appendix 1).

6. DISCUSSION

In the preceding section we have discussed in detail one of the possible resonant effects in magnetic-impurity scattering, guided primarily by what is new about the predicted phenomenon. Here we shall not begin to approach the specific implementation of an experiment, but must nonetheless mention the fundamentally nonequilibrium character of the phenomenon, which, in turn, should very likely necessitate the external excitation of donors. The natural means for this purpose is a radiation source operating in the submillimeter range. In germanium with $\mathbf{B}\|\langle 110\rangle$ the $1s(A_1)\to 2p_-$ transition (A_1 is a unitary representation of the lower split doublet; see Fig. 7) is an intravalley, i.e., ''vertical,'' transition, so that direct resonant excitation is possible with a frequency determined by the energy $E_{2p_-}-E_{1s(A_1)}$ in a field $B_0\approx 16.7$ T.

Vertical excitation from the ground state in Si with $\mathbf{B}\|\langle 001\rangle$ is allowed if the localized electron stays inside an E-valley; it is evident from Fig. 4 that elastic magnetic-

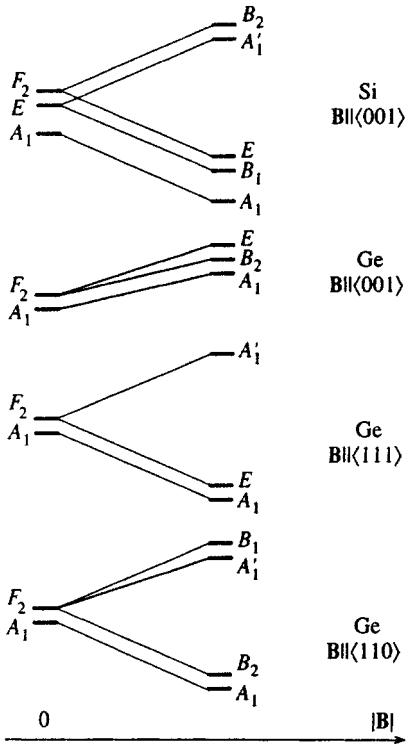

FIG. 7. Diagrams showing the lifting of magnetic degeneracy of the levels with a substitutional impurity in silicon and germanium. The letters to the left of the diagrams indicate the representations of the local symmetry group of the impurity in zero field, and the letters to the right indicate the representations of the local group in a finite field (see Table III).

impurity resonance must occur in fields not actually attainable under steady-state experimental conditions. The excitation $1s^{(E)}(A_1)\rightarrow 2p_{-1}^{(A)}$ in Si at $B_0\approx 10.5$ T is indirect. On the other hand, it is clear that under the given conditions in silicon at $T\gtrsim 1$ K the ''second ground state'' $1s^{(A)}(A_1)$ will be filled to a large extent even in equilibrium, because the energy difference between the $1s^{(A)}(A_1)$ and $1s^{(E)}(A_1)$ states is slight (see Figs. 5 and 7). Obviously, the chemical shift does not significantly alter this difference, and as a result external excitation to the $2p_{-1}^{(A)}$ state can be vertical with energy $E_{2p_{-1}^{(A)}}-E_{1s^{(A)}(A_1)}$.

We note that the strictly selective populating of resonant states is not necessarily guaranteed as a result of excitation. Nonresonant states (e.g., $2p_0$) are also going to be excited in some measure, but the efficiency with which they scatter is far lower in the absence of long-range influence (the dipole moment is equal to zero). Their scattering contribution can be estimated, once again with the aid of Eq. (5.23), on the assumption that n_d and n_d^* are the densities of nonresonance-excited and resonance-excited impurities. Here the contribution to the transverse conductivity is characterized by a ratio analogous to the right-hand side of (5.23); only the numerical factor changes, and Λ vanishes in the denominator. Of course, even though the estimate formally remains the same as before, the probability of scattering by an excited nonresonant donor is somewhat higher than in the ground state on account of the effective increase in a_B. At the same time, it is clear that this nonresonance scattering channel is not nearly as ''dangerous'' as the scattering by ground-state donors discussed in the derivation of (5.23), because the density of such donors is much greater than n_d^*.

We note, in addition, that scattering calculated by the Erginsoy model is too strong in comparison with the results of real experiments (see Ref. 2, Sec. 9.1, and Refs. 39 and 50). As a result, the nonresonant channels (scattering by donors in the ground or excited states) might well be even less significant than suggested by the estimates.

Another characteristic of experimental work is the need for free electrons in the material. In the given situation this condition is logically achieved by direct injection using ohmic contacts. The nonequilibrium of the carriers fosters the assumption that the excitation of donors is probably spontaneous for the most part (and, of course, nonselective), i.e., some of the donors will be excited as a result of the inelastic scattering of band electrons by them. The impurities can be distributed among their states in a substantially non-Gibbs fashion, and the ratio n_d^*/n_d can therefore be much greater than its equilibrium value. It is important to exclude ionization, i.e., the free-electron kinetic energy ϵ must still be lower than the donor binding energy in the ground state.

In closing, we call attention to the fact that the method developed in Secs. 2–4 enables us to predict and, in principle, to calculate the behavior of the transport characteristics for other cases of magnetic-impurity resonances. For example, let us consider Figs. 1 and 3. The main inelastic-scattering peak corresponds to coincidence of the intravalley excitation energy E_i-E_{1s} with the cyclotron energy, which is equal to $\hbar\Omega_B$ (Fig. 1) or to $\hbar\Omega_E=\hbar\Omega\sqrt{\gamma}$ (Fig. 3). In the effective-mass approximation the resonant values of the magnetic field for $i=2p_0, 2p_-, \ldots$ are easily determined by graphical means from the intersection of the dashed curves in Figs. 1 and 3 with the corresponding solid curves of the excitation spectrum. Such a resonance is impossible for Si, because the energy $E_{1s}+\hbar\Omega\sqrt{\gamma}$ is always much lower than E_i $(i\neq 1s)$ for real values of $|\mathbf{B}|$, even though in principle high-multiple resonances are possible when $E_i-E_{1s}=M\hbar\Omega\sqrt{\gamma}$, where $M\geqslant 4$.

Finally, we consider values of the field for which the edge of the conduction band (dot–dash lines in Figs. 1–3) crosses the energy level of the localized state. These crossings are probable points of resonance at the quasidiscrete level.[51] In principle, resonance of this type for semiconductors in a strong magnetic field has been investigated theoretically by Andreev,[53,54] who discusses a formulation of the problem requiring the satisfaction of certain conditions on the impurity scattering potential and the magnitude of the anisotropy. No light is shed on whether these conditions are achievable in reality. In our case resonant scattering at the quasidiscrete level is a natural outcome; the donor must be assumed to be ionized, rendering the experimental situation very simple. It follows from Fig. 2 that such a resonance can occur in Si only for states with $m>0$ (see the inset to Fig. 2), whereas in Ge it should occur and, in principle, be observable for all the investigated localized states, possibly with the exception of $2p_0^{(A)}$ (see Fig. 1). The resonant values of the magnetic field are determined fairly accurately from the figures by graphical means, where it is clear that the use of the

TABLE III.

	Field	Crystal group (magnetic class)	Local substitutional (impurity group)	State splitting (diagram)	Genesis of states for $\mathbf{B}\neq 0$ (Fig. 7)
Si	$\mathbf{B}=0$	O_h	T_d	A_1+E+F_2	
	$\mathbf{B}\parallel\langle 001\rangle$	$D_{4h}(C_{4h})$	D_{2d}	Doublet$\rightarrow A_1'+B_2$ Quadruplet $\downarrow$ A_1+B_1+E	$F_2\rightarrow B_2+E$ $E\rightarrow A_1'+B_1$
Ge	$\mathbf{B}=0$	O_h	T_d	A_1+F_2	
	$\mathbf{B}\parallel\langle 001\rangle$	$D_{4h}(C_{4h})$	D_{2d}	Quadruplet $\downarrow$ A_1+B_2+E	$F_2\rightarrow B_2+E$
	$\mathbf{B}\parallel\langle 111\rangle$	$D_{3d}(S_6)$	C_{3v}	Triplet$\rightarrow A_1+E$ Singlet$\rightarrow A_1'$	$F_2\rightarrow A_1'+E$
	$\mathbf{B}\parallel\langle 110\rangle$	$D_{2h}(C_{2h})$	C_{2v}	Lower doublet $\downarrow$ A_1+B_2 Upper doublet $\downarrow$ $A_1'+B_1$	$F_2\rightarrow A_1'+B_1+B_2$

effective-mass approximation is fully justified (the $1s$ state is not in any way involved in this effect). Calculations aimed at characterizing the resonances (see Figs. 1 and 2) must involve the small and in fact unknown matrix element of intervalley transition from the band formed by a type-B triplet in Ge or an E-quadruplet in Si to the bound state of an A-valley. On the other hand, it is clear that this fact is not a hindrance to obtaining the correct field dependence in the vicinity of resonance points in the final results; the indicated matrix element cannot harbor a dependence on **B**, because it is governed by a characteristic distance of the order of the lattice constant, which is much shorter than the magnetic length. Finally, the given type of resonance can be an intravalley resonance. Indeed, it follows from Fig. 3 that the level of the $2p_+^{(E)}$ states intersects the edge of the conduction band, whose position is determined by the very same E-valley.

The authors are grateful to S. I. Gubarev, V. N. Zverev, and S. V. Iordanskiĭ for a discussion, and also to I. A. Lebedev for assisting with some of the calculations. One of the authors (Dikman) is grateful for support from the Russian Fund for Fundamental Research (Grants 95-02-05883 and 96-02-17535).

APPENDIX 1

Table III gives the results of a group–theoretic analysis, which demonstrates how the valley–orbital degeneracy is lifted for S-states of a substitutional impurity in the presence of a finite magnetic field. We use standard notation for groups and their representations.[45] The last column of Table III and Fig. 7 show how the states in a finite field evolve from the states of the substitutional impurity at $\mathbf{B}=0$. The fact that the triplet in germanium splits into a singlet and a doublet for $\mathbf{B}\parallel\langle 001\rangle$ is ignored in Ref. 15. The splitting, of course, is slight by virtue of the smallness of the ratio of the lattice constant to the magnetic length.

APPENDIX 2

Here we discuss Eqs. (5.5), (5.14), (5.21), (5.26), (5.27′), (5.32), and (5.33′) in the case $\Delta E=\Delta\mathscr{E}=\zeta=0$. The integral in (5.21) and the representations (5.27) and (5.33) for the functions $G_{\parallel,\perp}^{A,E}$ reduce to the following expression after the integration of (5.17) over z:

$$I=\frac{4\varrho}{\pi^3}\int_0^\infty \beta(v)dv\ \exp\left(-\frac{|\eta|^2v^2}{2\varrho}\right)\int_{-\infty}^{\infty}du|Q(u,v)|^2. \tag{A2.1}$$

Here $\beta(v)$ is equal to the constant $2|\mathscr{D}|^2\pi^3$ or 1/4 if the integral (5.21) or (5.27) is calculated, or $\beta=v^2$ if the function in question is $G_\perp(\eta,0)$. The equation for $Q(u,v)$ has the form

$$Q=\int_{-\infty}^{\infty}dy\int_0^\infty dx\ \frac{\exp[-\varrho(y-u+iv\vartheta/2)^2]}{\sqrt{x^2+y^2}}\cos(vx), \tag{A2.2}$$

where the integral over x is tabulated and leads to the (modified Bessel function of the second kind) $K_0(|vy|)$. As a result, the integral of $|Q|^2$ is written in the form

$$\int_{-\infty}^{\infty}duQQ^*=\int_{-\infty}^{\infty}duR(v)\times\exp\left[\frac{\varrho\vartheta^2v^2}{2}-2\varrho\left(u-\frac{y_1+y_2}{2}\right)^2\right]$$

$$= \sqrt{\frac{\pi}{2\varrho}}\, R(v)\exp\left(\frac{\varrho\vartheta^2 v^2}{2}\right), \qquad \text{(A2.3)}$$

where

$$R(v)=\int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2 K_0(|vy_1|)K_0(|vy_2|) \times \exp\left[iv\vartheta\varrho(y_1-y_2)-\frac{\varrho(y_1-y_2)^2}{2}\right]. \qquad \text{(A2.4)}$$

We now substitute Eq. (A2.3) into (A2.1):

$$I=\frac{2}{\pi^3}\sqrt{2\pi\varrho}\int_0^{\infty}\beta(v)dv e^{-\varrho v^2/2}R(v). \qquad \text{(A2.5)}$$

We transform from the variables y_1 and y_2 in (A2.4) to $s=v(y_1+y_2)/2$ and $r=y_1-y_2$:

$$R(v)=\frac{1}{v}\int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} dr K_0\left(\left|s+\frac{vr}{2}\right|\right)K_0\left(\left|s-\frac{vr}{2}\right|\right) \times \exp\left(-\frac{\varrho r^2}{2}+i\varrho\vartheta v r\right). \qquad \text{(A2.6)}$$

Consequently, if $\beta(v)=\text{const}$, the integral over v (A2.5) diverges logarithmically to zero. In the presence of a cutoff mechanism at small $v\sim\overline{v}$ it can be evaluated, in which case the integrand of Eq. (A2.6) is large for $s\sim r\sim 1\gg vr$, so that

$$R(v)\approx\frac{1}{v}\sqrt{\frac{2\pi}{\varrho}}\int_{-\infty}^{\infty}K_0^2(|s|)ds. \qquad \text{(A2.7)}$$

The integral in this equation is computed analytically (see Ref. 54, Sec. 2.16.33.2) and is equal to $\pi^2/2$. Thus, for $\beta(v)=2|\mathscr{D}|^2\pi^3$ we have in (5.21)

$$I=\int dv\mathscr{S}(v,0)\Big|_{\Delta E=0}=4\pi^3|\mathscr{D}|^2\int_{v>|\overline{v}|}\frac{dv}{v}. \qquad \text{(A2.8)}$$

We now simplify Eq. (2.5) in the case $\beta=v^2$. Transforming to the variables $t_1=vy_1$ and $t_2=vy_2$ in (A2.6), we substitute (A2.6) into (A2.5) and carry out the tabulated integration over v. Hence,

$$G_{\perp}(\eta,0)=\frac{2}{\pi^2}\int_{-\infty}^{\infty}dt_1\int_{-\infty}^{\infty}dt_2K_0(|t_1|)K_0(|t_2|) \times\exp[-\varrho|t_1-t_2|+i\varrho\vartheta(t_1-t_2)]. \qquad \text{(A2.9)}$$

If $\varrho\ll 1$ or if $\varrho\gg 1$, then for any values of $\varrho\vartheta$ the integral (A2.9) can be factored in leading order. In the first case it reduces to $(\int K_0(t)\cos(\varrho\vartheta t)dt)^2$, and in the second case it reduces to the integral already analyzed above, $\int K_0(|t|)^2dt$. Therefore, taking the expression (5.10) for $\eta(\alpha)$ into account, we find

$$G_{\perp}(\eta,0)=\begin{cases}2\sin^2\alpha & \text{if } \gamma\ll 1, \text{ but } \alpha\gg\sqrt{\gamma},\\ 2\sqrt{\gamma}, & \text{if } \alpha^2\lesssim\gamma\ll 1.\end{cases} \qquad \text{(A2.10)}$$

Finally, the integral (A2.9) is evaluated exactly in the case $\eta\equiv 1$ (i.e., $\varrho=1$, $\vartheta=0$). It is readily inferred with the use of tables (see Ref. 54, Secs. 1.12.2.2 and 1.12.2.3) that

$$G_{\perp}(1,0)\equiv 1. \qquad \text{(A2.11)}$$

APPENDIX 3

The justifiability of the Born approximation depends significantly on precisely which matrix elements $\langle\alpha|U|\nu\rangle$ determine the principal contribution to the investigated transport scattering characteristics. Here again the $|\nu\rangle$ state is the ''unperturbed'' wave function (5.9), and the $\langle\alpha|$ state is generally interpreted as an exact function of the continuous spectrum Φ_α in the potential U, corresponding to the energy $\hbar\Omega_z/2+\epsilon$. The Born approximation is strictly valid if in the range of r significant for the calculation of the characteristic matrix elements it can be assumed that $\Phi_\alpha\approx\Psi_\alpha$. On the other hand, this condition is often violated (e.g., in the problem of scattering by charged particles subject to the condition $a_B\lesssim\lambda$; see Ref. 2), even though the relation $\Phi_\alpha\sim\Psi_\alpha$ holds in the characteristic range. The Born approximation then gives a semiquantitative result[2] and is still suitable for estimates. This is in fact the situation in the calculation of the transverse conductivity in our case. Indeed, we have seen that when the function (5.33) is reduced to dimensionless form, it is of order unity if $\mathscr{D}\sim 1$. This means that the interaction zone of an electron with an impurity center, being the principal contributing factor to the final result, is of order λ, and the characteristic scattering potential (5.5) is of order $\hbar\Omega$. Accordingly, the electron wave function in the interaction zone is distorted, but is still of the same order of magnitude as the unperturbed wave function. On the other hand, if $\mathscr{D}$ is substantially smaller than unity, our result is completely rigorous in the sense of the validity of the Born approximation. In any case it reflects the functional dependence of the transverse conductivity on the magnetic field.

We now consider the calculation of the collision frequency (5.15) and the longitudinal conductivity. Here the conditions for validity of the Born approximation are strictly satisfied in the vicinity of resonance. For the estimation of Φ_α it is convenient to use the exact equation

$$\Phi_\alpha(\mathbf{r})=\Psi_\alpha(\mathbf{r})+\int\mathscr{G}(\epsilon,\mathbf{r},\mathbf{r}')U(\mathbf{r}')\Phi_\alpha(\mathbf{r}')d\mathbf{r}' \qquad \text{(A3.1)}$$

(see, e.g., Ref. 55). In our case of a quantizing magnetic field the Green's function $\mathscr{G}(\epsilon,\mathbf{r},\mathbf{r}')$ is almost independent of z and z', is a maximum for $|[(\mathbf{r}-\mathbf{r}')\cdot\mathbf{B}]|/B\lesssim\lambda$, and decays rapidly like the exponential of a quadratic outside this range. At the maximum the function $\mathscr{G}$ is of order magnitude $1/\lambda^3\sqrt{\epsilon\hbar\Omega}$, so that the correction to Ψ_α is of order $U(\mathbf{r})/\sqrt{\epsilon\hbar\Omega}$. We have seen that in the calculation of τ^{-1} the characteristic value of U is determined by the distance $r\sim x\sim\lambda\sqrt{\hbar\Omega/\delta E}$, so that the integral (5.21), which diverges for small values of v, can be cut off. We therefore have the characteristic value $U(\mathbf{r})\sim\hbar\Omega\lambda^2/r^2\sim\delta E$. Assuming $\delta E\sim\epsilon$, we obtain a correction, small in proportion to the parameter $\sqrt{\epsilon/\hbar\Omega}$, to the wave function in the interaction zone.

*E-mail: dickmann@issp.ac.ru

[1)]This method has been applied to the donor problem in Ref. 30.

[2)]For effective masses we use the values $m_e/m_z=0.0816$ and 0.1905, $\gamma=0.05134$ and 0.208 for germanium and silicon, respectively (see Ref. 40), and $\kappa=15.4$ and 11.4 found on the basis of a comparison with optical measurements (see Refs. 31, 33, and 41). Note that these values of κ are substantially lower than those measured by the capacitance method.[42]

[3)]In this case the discussion concerns the magnetic group, i.e., allowance is made for the fact that the axial vector **B**, directed along the C_∞ axis, changes sign in the presence of rotations u_2 and reflections σ_d. The vector **B** is invariant under all other representations of this group, including the reflection σ_h. But if we turn directly to the Schrödinger equation (2.1), the field enters into it simply as a numerical parameter $\mathscr{B}$, so that the symmetry group of this equation is $C_{\infty h}$. To avoid misunderstanding, we also note that the solutions of the Schrödinger equation which depend on the angle φ are the basis functions of representations of the group $C_{\infty h}$. In order for them to be preserved as basis functions for representations of the group $D_{\infty d}$, the magnetic–field dependence must be included explicitly in $\exp(im\varphi)$ by making the substitution $\varphi\to\varphi\mathbf{n}_z\cdot\mathbf{B}/|\mathbf{B}|$, where $\mathbf{n}_z$ is the unit vector in the positive z direction.

[4)]This result is itself an indirect confirmation of the ''correctness'' of the resulting wave functions in states 1 and 2.

[5)]By virtue of (5.1), the role of an external electric field in lifting the degeneracy of the $2s-2p_-$ state is even more negligible by comparison when the field of the band electron is taken into account.

[6)]If $B=0$ holds, then τ_{Er} does not depend on the carrier energy in the model of Ref. 39.

[1] V. F. Gantmakher and V. N. Zverev, in *Landau Level Spectroscopy*, edited by G. Landwehr and E. I. Rashba, (North-Holland, Amsterdam, 1991), p. 1135.

[2] V. F. Gantmakher and I. B. Levinson, *Carrier Scattering in Metals and Semiconductors* (North-Holland, Amsterdam, 1987).

[3] L. Eaves and J. C. Portal, J. Phys. C **12**, 2809 (1979).

[4] V. F. Gantmakher and V. N. Zverev, Zh. Éksp. Teor. Fiz. **69**, 695 (1975) [Sov. Phys. JETP **42**, 352 (1975)].

[5] V. N. Zverev and D. V. Shovkun, Zh. Éksp. Teor. Fiz. **87**, 1745 (1984) [Sov. Phys. JETP **60**, 1003 (1984)]; Fiz. Tverd. Tela (Leningrad) **26**, 356 (1984) [Sov. Phys. Solid State **26**, 212 (1984)].

[6] K. Von Klitzing and J. Tuchendler, *Physics in High Magnetic Fields*, edited by S. Chikazumi and N. Miura (Springer-Verlag, Berlin-New York, 1981), p. 139.

[7] V. L. Gurevich, JETP Lett. **5**, 210 (1967).

[8] S. M. Dikman and D. I. Sidel'nikov, JETP Lett. **53**, 602 (1991).

[9] S. M. Dickmann and D. I. Sidel'nikov, Phys. Lett. A **187**, 79 (1994).

[10] B. Pajot, F. Merlet, and G. Taravella, Can. J. Phys. **50**, 2186 (1972).

[11] Y. M. Mu, J. P. Peng, P. L. Liu, S. C. Shen, and J. P. Zhu, Phys. Rev. B **48**, 10864 (1993).

[12] S. C. Shen, Solid State Commun. **93**, 357 (1995).

[13] E. M. Gershenzon, G. N. Gol'tsman, and A. I. Elant'ev, Zh. Éksp. Teor. Fiz. **72**, 1062 (1977) [Sov. Phys. JETP **45**, 555 (1977)].

[14] B. Pajot, F. Merlet, G. Taravella, and P. Arcas, Can. J. Phys. **50**, 1106 (1972).

[15] N. Lee, D. M. Larsen, and B. Lax, J. Phys. Chem. Solids **34**, 1817 (1973).

[16] N. Lee, D. M. Larsen, and B. Lax, J. Phys. Chem. Solids **35**, 401 (1974).

[17] D. Cabib, E. Fabrio, and G. Fiorio, Nuovo Cimento B **10**, 185 (1972).

[18] M. S. Kashiev, S. I. Vinitsky, and F. R. Vukajlovic, Phys. Rev. A **22**, 557 (1980).

[19] J. M. Wadehra, Astrophys. J. **248**, 874 (1981).

[20] A. V. Turbiner, Usp. Fiz. Nauk **144**, 35 (1984) [Sov. Phys. Usp. **27**, 668 (1984)].

[21] A. V. Turbiner, Yad. Fiz. **46**, 204 (1987) [Sov. J. Nucl. Phys. **46**, 125 (1987)].

[22] A. V. Turbiner, J. Phys. A **17**, 859 (1984).

[23] R. J. Elliott and R. Loudon, J. Phys. Chem. Solids **8**, 382 (1959); J. Phys. Chem. Solids **15**, 196 (1960).

[24] W. Rösner, H. Herold, H. Ruder, and G. Wunner, Phys. Rev. A **28**, 2071 (1983).

[25] G. Wunner, W. Rösner, H. Ruder, and H. Herold, Astrophys. J. **262**, 407 (1982).

[26] W. Rösner, G. Wunner, H. Herold, and H. Ruder, J. Phys. B **17**, 29 (1984).

[27] H. Forster, W. Strupat, W. Rösner, G. Wunner, H. Ruder, and H. Herold, J. Phys. B **17**, 29 (1984).

[28] M. V. Ivanov, J. Phys. B **17**, 29 (1984).

[29] Yu. P. Kravchenko, M. A. Liberman, and B. Johansson, Phys. Rev. A **53**, 287 (1996); Phys. Rev. Lett. **77**, 619 (1996).

[30] W. Kohn and J. M. Luttinger, Phys. Rev. **97**, 883 (1966); Phys. Rev. **98**, 915 (1955).

[31] R. A. Faulkner, Phys. Rev. **184**, 713 (1969).

[32] J. Broeckz, P. Claws, and J. Vennik, J. Phys. C **19**, 511 (1986).

[33] I. L. Beĭnikhes and Sh. M. Kogan, Zh. Éksp. Teor. Fiz. **93**, 285 (1987) [Sov. Phys. JETP **66**, 164 (1987)].

[34] V. I. Galiev, A. F. Polupanov, and I. E. Shparlinski, J. Comput. Appl. Math. **39**, 151 (1992).

[35] E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Pergamon Press, New York–Oxford, 1981) [Russ. original, Nauka, Moscow, 1979].

[36] S. Titejca, Ann. Phys. (Leipzig) **22**, 128 (1935).

[37] B. I. Davydov and I. Ya. Pomeranchuk, Zh. Éksp. Teor. Fiz. **9**, 1294 (1939).

[38] E. Adams and T. Holstein, J. Phys. Chem. Solids **10**, 254 (1959).

[39] C. Erginsoy, Phys. Rev. **79**, 1013 (1950).

[40] *Landolt–Börnstein*, Vol. 17a (Springer-Verlag, Berlin, 1982); *Landolt–Börnstein*, Vol. 22a (Springer-Verlag, Berlin, 1987).

[41] I. L. Beĭnikhes and Sh. M. Kogan, JETP Lett. **44**, 48 (1986).

[42] K. V. Rao and A. Smakula, J. Appl. Phys. **37**, 2840 (1966).

[43] N. Lee, D. M. Larsen, and B. Lax, J. Phys. Chem. Solids **34**, 1059 (1973).

[44] M. M. Vaĭnberg, *Variational Method and Method of Monotonic Operators in the Theory of Linear Equations* [in Russian] (Nauka, Moscow, 1972).

[45] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Israel Program for Scientific Translations, Jerusalem; Wiley, New York, 1975).

[46] A. K. Ramdas and S. Rodriguez, Rep. Prog. Phys. **44**, 1297 (1981).

[47] C. Jagannath, Z. W. Grabowski, and A. K. Ramdas, Phys. Rev. B **24**, 2082 (1981).

[48] C. Jacoboni, C. Canali, G. Ottaviani, and A. Quaranta, Solid-State Electron. **20**, 77 (1977).

[49] C. Jacoboni, F. Nava, C. Canali, and G. Ottaviani, Phys. Rev. B **24**, 1014 (1981).

[50] H. Kobori, T. Ohyama, and E. Otsuka, J. Phys. Soc. Jpn. **59**, 2164 (1990).

[51] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon Press, New York–Oxford, 1974) [Russ. transl., Nauka, Moscow, 1974].

[52] S. P. Andreev, Zh. Éksp. Teor. Fiz. **77**, 1046 (1979) [Sov. Phys. JETP **50**, 526 (1979)].

[53] S. P. Andreev, Usp. Fiz. Nauk **143**, 213 (1984) [Sov. Phys. Usp. **27**, 431 (1984)].

[54] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series*, Vols. 1–3, Gordon and Breach, New York, 1986, 1986, 1989) [Russ. original, Vols. 1–3, Nauka, Moscow, 1981, 1983, 1986].

[55] V. G. Skobov, Zh. Éksp. Teor. Fiz. **37**, 1467 (1959) [Sov. Phys. JETP **10**, 1039 (1959)].

Translated by James S. Wood

Possibility of observing magnetic macroscopic quantum tunneling using a scanning tunneling microscope

S. N. Molotkov

Institute of Solid-State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia
(Submitted 8 October 1996)
Zh. Éksp. Teor. Fiz. **112**, 1011–1020 (September 1997)

A theory describing the detection of magnetic macroscopic quantum tunneling on the basis of current fluctuations in a scanning tunneling microscope is proposed. © *1997 American Institute of Physics.* [S1063-7761(97)01609-0]

1. INTRODUCTION

The idea that magnetic macroscopic quantum tunneling is possible in tiny magnetic particles was probably first advanced in Ref. 1, although the question was raised in a very general form back at the dawn of the creation of quantum mechanics and was formulated by Schrödinger in the form of a familiar paradox known as the paradox of Schrödinger's cat,[2] which reduces, in effect, to the following statement. If quantum mechanics is correct, it should be applicable to the description of macroscopic objects, and if this is, in fact, so, the superposition of different states of a macroscopic system should take place. The question of observing such a superposition remained open for a long time. Magnetic nanoparticles are objects, which are simultaneously macroscopic and still small enough that quantum-mechanical phenomena, such as quantum tunneling, could be exhibited in them.

Small magnetic particles have served as objects for observing macroscopic quantum tunneling.[3]

Quantum tunneling is a transition of a microscopic system from one state to another along a classically forbidden trajectory. If the system was created in one of these states (and weak tunneling transitions are disregarded), the system is found to be in a superposition of different macroscopic states when the transitions to other states are taken into account. Such transitions are essentially ordinary quantum beats, which should be manifested experimentally in the form of a low-frequency collective mode.

Following the publication of Ref. 3 there were numerous papers devoted to magnetic macroscopic quantum tunneling.[4–11] Macroscopic quantum tunneling was investigated experimentally by performing low-temperature magnetic relaxation measurements in Ref. 10 and 11, as well as by measuring the response to a small magnetic field using a SQUID microsusceptometer in Ref. 9.

In this paper we wish to focus attention on one possibility of detecting macroscopic quantum tunneling using a scanning tunneling microscope. This instrument has been used successfully to detect the local magnetization on a surface with atomic resolution.[12] The frequency spectrum of the noise (fluctuations) in the tunneling current from the surface of a Si(111)(7×7) sample immersed in a constant magnetic field ($H\approx100$ G) was investigated in Ref. 13. The surface was light bombardment preliminarily subjected to by oxygen, which resulted in the formation of paramagnetic centers on the surface. The peak discovered in Ref. 13 in the tunneling current noise spectrum at the Larmor frequency ($\Omega\approx480$ MHz) exceeded the background noise in intensity by 16 dB. Moving the scanning tunneling microscope tip over the surface to a distance of several angstroms ($\sim3-5$ Å) from the paramagnetic center led to disappearance of the signal.[13] Manassen *et al.* interpreted this peak as a result of the Larmor precession of the spin of an individual electron localized on a paramagnetic center. A theory of the possible mechanisms responsible for this feature was presented in Ref. 14. The temporal dependence (rather than the noise spectrum) of the tunneling current for tunneling from a metal substrate into a scanning microscope tip through a large organic molecule was investigated directly in a somewhat later experiment.[15] The configuration of the magnetic fields in that experiment corresponded to the standard configuration for electron paramagnetic resonance (EPR) investigations, i.e., a constant magnetic field and a weak rotating magnetic field perpendicular to it. Temporal oscillations of the tunneling current were discovered at a frequency corresponding to the EPR frequency for an electron on a free radical in the molecule. Moving the microscope tip over the surface to a distance of several angstroms from the site of the radical caused the signal to disappear. A theory of EPR in a scanning tunneling microscope was devised in Ref. 16. The appearance of a spin-dependent component in the current was also predicted for tunneling from a GaAs surface irradiated by circularly polarized light.[17] The appearance of such a component in the current was detected experimentally in Ref. 18.

Thus, since a scanning tunneling microscope is a local probe, it is capable, in principle, of detecting fluctuations of individual spins in a tiny magnetic particle.

In this paper we wish to demonstrate that a peak should be observed in the frequency spectrum of the tunneling current fluctuations at a frequency corresponding to the low-frequency mode associated with macroscopic quantum tunneling in the case of tunneling through a tiny magnetic particle on a surface into the microscope tip.

The experimental observation of magnetic macroscopic quantum tunneling using a scanning tunneling microscope can be simpler than the observation of the effects mentioned above, since the characteristic frequencies of macroscopic quantum tunneling can be significantly lower than the frequencies at which EPR is observed. In addition, since each

spin makes an additive contribution, the amplitude of the signal should be considerably greater than the signal from an individual electron spin. It is also significant that, in contrast to the experiments in Refs. 13 and 15, the observation of macroscopic quantum tunneling by a scanning tunneling microscope does not require an external magnetic field.

2. BASIC EQUATIONS

Let us consider a system consisting of a metal substrate (M) with a magnetic particle (MP) on it under a scanning tunneling microscope tip (T) (Fig. 1). We shall consider a system consisting of a quantum spin **S**, which can undergo quantum tunneling, as a model for describing a tiny magnetic particle. The Hamiltonian describing the magnetic particle will not be specified. It will be sufficient for us to know only the frequency-dependent behavior of the spin–spin correlator for the magnetic particle, which contains information on quantum tunneling in an isolated particle. It is assumed that the electron–electron interaction ''freezes'' the total spin in the particle, and the magnetization vector has at least two states of equivalent energy (if the possible tunneling between them is disregarded), whose positions are found as classical energy minima. Quantum tunneling is described as a subbarrier (tunneling) transition of the magnetization vector **S** between two orientations corresponding to different energy minima.

When tunneling occurs through a magnetic particle, we assume that there is a level for an extra electron in it. Such an assumption is quite natural. Next, in order to simplify the mathematical manipulations, we assume that the particle is a point (one site with a spin). In a real situation the spins are localized in a particle on separate sites, each of which is coupled with atoms in the tip and the substrate by a tunneling matrix element. Such a generalization creates technical complications, but does not influence the qualitative result. We shall discuss this generalization later on. The electron states in the tip and the substrate will be described as single-particle states.

We write the Hamiltonian of the system in the form

$$\begin{aligned}\hat{H}&=\hat{H}_a+\hat{H}_b+\hat{H}_{\text{spin-STM}}+\hat{T}_a+\hat{T}_b+\hat{H}_{\text{spin}}(\hat{\mathbf{S}})\\&=\sum_{k,\sigma}(\varepsilon_{ka}-\mu_t)a^+_{k\sigma}a_{k\sigma}+\sum_{k,\sigma}(\varepsilon_{kb}-\mu_c)b^+_{k\sigma}b_{k\sigma}\\&\quad+\sum_{\sigma,\sigma'}Jc^+_\sigma c_{\sigma'}S_{\sigma\sigma'}+\sum_\sigma\varepsilon_0 c^+_\sigma c_\sigma\\&\quad+\sum_k(T_{ka}c^+_\sigma a_{k\sigma}+T_{kb}c^+_\sigma b_{k\sigma}+\text{H.c.})+\hat{H}_{\text{spin}}(\hat{\mathbf{S}}).\end{aligned}\tag{1}$$

The first two terms in (1) describe the spectrum of the electrons in the substrate and the tip. The next term corresponds to the interaction of the spin of a tunneling electron and the spin in the particle, the fourth term specifies the state of the tunneling electron at the level for the extra electron in the particle, and the fifth term defines the tunneling coupling between the substrate and the particle and between the particle and the tip. The last term corresponds to the Hamiltonian of the isolated magnetic particle. The operator of the

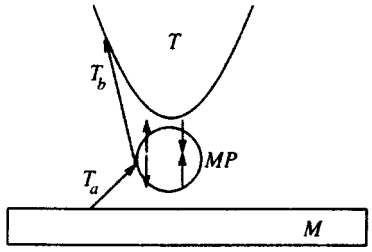

FIG. 1. Schematic representation of the experiment.

tunneling current from the substrate to the tip through the particle is described in the usual manner,[19] and in the Heisenberg representation it has the form

$$\begin{aligned}\hat{I}(t)&=\hat{I}_a(t)+\hat{I}_b(t),\\\hat{I}(t)&=\frac{ie}{2}\left\{\sum_{k\sigma}[T_{ka}c^+_\sigma(t)a_{k\sigma}(t)-T^*_{ka}a^+_{k\sigma}c_\sigma(t)]\right.\\&\quad\left.+[T_{kb}b^+_{k\sigma}(t)c_\sigma(t)-T^*_{kb}c^+_\sigma b_{k\sigma}(t)]\right\}.\end{aligned}\tag{2}$$

Our ultimate purpose is to calculate the current–current correlator, since this is the quantity which is measured experimentally when precession of the spin of an individual electron is detected[13] (in Ref. 13 this quantity was called the radio-frequency power spectrum). The current–current correlator is written in the usual manner:

$$\begin{aligned}K(t-t')&=\langle[\hat{I}(t)\hat{I}(t')+\hat{I}(t')\hat{I}(t)]\rangle_K=\langle[\hat{I}(t)\hat{I}(t')\\&\quad+\hat{I}(t')\hat{I}(t)]\hat{S}\rangle,\end{aligned}\tag{3}$$

where the subscript K indicates that the averaging is carried out over a Keldysh contour, and the $\hat{S}$ matrix has the form

$$\begin{aligned}\hat{S}(-\infty,\infty)&=\hat{T}_K\exp\left\{i\int_{-\infty}^{\infty}[\hat{T}_a(\tau)+\hat{T}_b(\tau)\right.\\&\quad\left.+\hat{H}_{\text{spin-STM}}(\tau)]d\tau\right\},\end{aligned}\tag{4}$$

where $\hat{T}_K$ is the operator for chronological ordering over the Keldysh contour. To isolate the terms that interest us in the current–current correlator (3), we write it out in greater detail, and with consideration (2) we find

$$\begin{aligned}K(t-t')&=\left(\frac{e}{2}\right)^2\left\langle\{\hat{T}_a\hat{T}_b\hat{c}^+(t)\hat{a}(t)\hat{b}^+(t')\hat{c}(t')\right.\\&\quad\left.-\hat{T}_a\hat{T}^+_b\hat{c}^+(t)\hat{a}(t)\hat{c}^+(t')\hat{b}(t')\}\hat{S}(-\infty,\infty)\right.\\&\quad-\left(\frac{e}{2}\right)^2\langle\{\hat{T}^+_a\hat{T}_b\hat{a}^+(t)\hat{c}(t)\hat{b}^+(t')\hat{c}(t')\\&\quad\left.-\hat{T}^+_a\hat{T}^+_b\hat{a}^+(t)\hat{c}(t)\hat{c}^+(t')\hat{b}(t')\}\hat{S}(-\infty,\infty)\right\rangle,\end{aligned}\tag{5}$$

where, for brevity, we have introduced the notation $\hat{T}_{a,b}=\{T_{ka,b}\}$, $\hat{c}=\{c_\sigma\}$, $\hat{a}=\{a_{k\sigma}\}$, $\hat{b}=\{b_{k\sigma}\}$, and $\hat{S}=\{S_{\sigma\sigma'}\}$ and where the corresponding summation must be carried out over the spin and the index k.

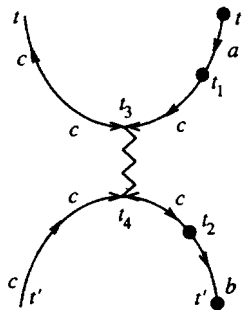

FIG. 2.

Expanding the $\hat{S}$ matrix into a series in $\hat{T}_{a,b}$ and $\hat{H}_{\text{spin-STM}}$ and grouping the terms appropriately, we can isolate the special terms. For example, the last term in Eq. (5) gives

$$\hat{T}_a\hat{T}_b^+\hat{T}_a^+\hat{T}_b J^2\int\int\int\int d\tau_1 d\tau_2 d\tau_3 d\tau_4$$
$$\times\langle[\hat{a}^+(t)\hat{c}(t)\hat{c}^+(t')\hat{b}(t')\hat{c}^+(\tau_1)\hat{a}(\tau_1)\hat{b}^+(\tau_2)\hat{c}(\tau_2)$$
$$\times\hat{c}^+(\tau_3)\hat{c}(\tau_3)\hat{c}^+(\tau_4)\hat{c}(\tau_4)\hat{S}(\tau_3)\hat{S}(\tau_4)]\rangle. \tag{6}$$

The diagram in Fig. 2 corresponds to this term and terms similar to it. The solid lines correspond to electron propagators, the filled circles correspond to the tunneling matrix elements T_a and T_b, the cross corresponds to the exchange interaction J, and the wavy line corresponds to the spin propagator in the magnetic particle.

The current–current correlator is obtained by closing the ends to the two-particle Green's function (Fig. 2) for the times t and t'. A total of eight similar diagrams containing a feature at the macroscopic quantum tunneling frequency appear. The diagrams for the current–current correlator obtained after taking the Fourier transform with respect to the time difference $t-t'$ are presented in Fig. 3. The external frequency Ω is indicated by a dashed line by convention. The structure of the diagrams and the correlator can be represented in the form

$$K(\Omega)\propto\Pi_a(\Omega)\langle\hat{S}\hat{S}\rangle_\Omega\Pi_b(\Omega), \tag{7}$$

where Π_a and Π_b are the polarization operators corresponding to the electron loops in Fig. 3. The contribution creating a feature arises from the spin–spin correlator in the magnetic particle at the frequency corresponding to magnetic macroscopic quantum tunneling. Due to the presence of a low-frequency macroscopic quantum tunneling (MQT) mode, the correlator $\langle\hat{S}\hat{S}\rangle_\Omega$ behaves as $\delta(\Omega-\Omega_{\text{MQT}})$.[4,8,9] Before proceeding to the detailed calculation, let us qualitatively discuss the reason for the appearance of the feature. The diagram in Fig. 2 describes the tunneling current fluctuations. The upper half of the diagram describes tunneling from the substrate into the magnetic particle followed by an interaction of the tunneling electron with the spin of the particle and its subsequent return to the substrate. Such passage of an electron excites a macroscopic quantum tunneling mode. A second electron from the tip passes into the particle and interacts with the already excited macroscopic quantum tunneling mode and then returns to the tip. Such processes describe the coherent tunneling of electrons from the substrate and the tip into the magnetic particle and their interaction with the macroscopic quantum tunneling mode. If a voltage is applied between the tip and the substrate, these processes are superposed on a constant mean tunneling current. In the absence of a voltage the mean tunneling current is equal to zero, and the processes appear as equilibrium current fluctuations.

The Fourier transform of the current–current correlator can be represented in the following form

$$\langle II\rangle_\Omega = K^{+-}(\Omega)+K^{-+}(\Omega),$$
$$K^{\pm\mp}(\Omega)=\left(\frac{e}{2}\right)^2\{(\Pi_{1a}(\Omega)-\Pi_{2a}(\Omega))\Gamma(\Omega)(\Pi_b(\Omega)$$
$$-\Pi_{2b}(\Omega))\}^{\pm\mp}+\left(\frac{e}{2}\right)^2\{(\Pi_{1b}(\Omega)$$
$$-\Pi_{2b}(\Omega))\Gamma(\Omega)(\Pi_a(\Omega)-\Pi_{2a}(\Omega))\}^{\pm\mp}, \tag{8}$$

where the superscripts $-+$ and $+-$ denote the corresponding Keldysh components. The polarization operators are defined in the following manner:

$$\Pi_{1a}(\Omega)=\int\hat{G}_{cc}(\Omega+\omega)\hat{T}_a\hat{G}_{ac}(\omega)\frac{d\omega}{2\pi},$$
$$\Pi_{1b}(\Omega)=\int\hat{G}_{cc}(\Omega+\omega)\hat{T}_b\hat{G}_{bc}(\omega)\frac{d\omega}{2\pi}, \tag{9}$$
$$\Pi_{2a}(\Omega)=\int\hat{G}_{cc}(\omega)\hat{T}_a\hat{G}_{ac}(\Omega+\omega)\frac{d\omega}{2\pi},$$
$$\Pi_{2b}(\Omega)=\int\hat{G}_{cc}(\omega)\hat{T}_a\hat{G}_{bc}(\Omega+\omega)\frac{d\omega}{2\pi},$$
$$\Gamma(\Omega)=J^2\langle\hat{S}\hat{S}\rangle_\Omega.$$

For brevity, in (8) and (9) we used notation similar to (5), which signifies summation over the spins. We shall turn to the definition of the Green's functions themselves later on. Each triple product in (8) should be expressed in terms of advanced, retarded, and Keldysh Green's functions. To accomplish this, it is useful to take advantage of the following property (see, for example, Ref. 19):

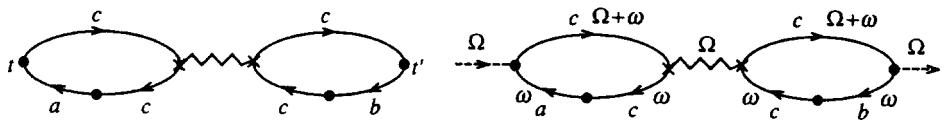

FIG. 3.

$$\{ABC\}^{+-}+\{CBA\}^{+-}+\{ABC\}^{-+}+\{CBA\}^{-+}$$
$$=(A^{+-}+A^{-+})(B^RC^R+C^AB^A)+(C^AA^R+A^AC^R)$$
$$\times(B^{+-}+B^{-+})+(A^AB^A+B^RC^R)$$
$$\times(C^{+-}+C^{-+}), \tag{10}$$

where A, B, and C denote one of the functions ($\Pi_{1a,2a}$, $\Pi_{1b,2b}$, and Γ). We have also introduced the notation $A=\Pi_{1a}-\Pi_{2a}$, $B=\Gamma$, and $C=\Pi_{1b-\Pi 2b}$. Information on the spectrum of macroscopic quantum tunneling fluctuations is contained in the third term in Eq. (10). It has a spike at Ω_{MQT}, since the Keldysh components $\Gamma^{+-}(\Omega)$ and $\Gamma^{-+}(\Omega)$ contribute to the generalized susceptibility appearing in the fluctuation–dissipation theorem:

$$\langle S_zS_z\rangle_\Omega=\Gamma^{+-}(\Omega)+\Gamma^{-+}(\Omega)=J^2\ \mathrm{const}\coth\left(\frac{\Omega}{2T}\right)$$
$$\times\delta(\Omega-\Omega_{\mathrm{MQT}}). \tag{11}$$

The calculation of the correlator $\langle S_zS_z\rangle_\Omega$ is a separate problem (see, for example, Refs. 4, 8, and 9). The peak in the fluctuation spectrum is a manifestation of the existence of a low-frequency feature associated with a collective macroscopic quantum tunneling mode.

The only remaining task is to calculate the advanced and retarded components of the electronic polarization operators, which, for example, for Π_{1a} can be brought into the form

$$\Pi^A_{1a}(\Omega)=\Pi^{--}_{1a}(\Omega)-\Pi^{-+}_{1a}(\Omega)=\hat T_a$$
$$\times\int\hat G^R_{cc}(\Omega+\omega)\hat T_a\hat G^A_{ac}\frac{d\omega}{2\pi}+\hat T_a$$
$$\times\int\hat G^R_{cc}(\Omega+\omega)\hat T_a\hat G^{+-}_{ac}\frac{d\omega}{2\pi}+\hat T_a$$
$$\times\int\hat G^{-+}_{cc}(\Omega+\omega)\hat T_a\hat G^A_{ac}\frac{d\omega}{2\pi}. \tag{12}$$

The polarization operator $\Pi_{1a}(\Omega)-\Pi_{2a}(\Omega)$ can be brought into the form

$$\Pi_{1a}(\Omega)-\Pi_{2a}(\Omega)$$
$$=\int[\hat G^R_{cc}(\Omega+\omega)\hat T_a\hat G^A_{ac}-\hat G^A_{cc}(\Omega+\omega)\hat T_a\hat G^R_{ac}]$$
$$=\sum_k\int|T_{ak}|^2G^R_{ak}(\omega)G^R_{cc}(\omega)[G^A_{cc}(\Omega+\omega)$$
$$-G^A_{cc}(\omega-\Omega)]\frac{d\omega}{2\pi}=N_a\langle|T_{ak_F}|^2\rangle\frac{2\Omega}{\Omega^2+(2\gamma)^2}, \tag{13}$$

where the following expressions are used for the Green's functions:

$$G^{R,A}_{ak}=\frac{1}{\omega-\varepsilon_{ak}\pm i0},\quad G^{R,A}_{cc}=\frac{1}{\omega-\varepsilon_c\pm i\gamma}. \tag{14}$$

In addition, it is assumed that the density of states in the substrate N_a is constant in the vicinity of the Fermi level (constancy is required only on the scale $\Omega_{\mathrm{MQT}}\ll\varepsilon_F$), and $\langle|T_{ak_F}|^2\rangle$ is the tunneling matrix element at the Fermi energy averaged over the angles. The Green's function for an electron at the level in the particle can be obtained in a standard way, as is done in the problem of tunneling between two massive reservoirs weakly coupled to one another through a level in a quantum well (see, for example, Ref. 20). In analogy to Ref. 20, for the width γ of the level in (14) we find

$$\gamma=\gamma_a+\gamma_b=\sum_k[|T_{ak}|^2\delta(\varepsilon_c-\varepsilon_{ak})$$
$$+|T_{bk}|^2\delta(\varepsilon_c-\varepsilon_{bk})]. \tag{15}$$

If the density of states in the reservoirs varies weakly in the energy range that interests us, γ does not depend on the frequency.

Next, the last term in (12) can be brought into the form

$$\hat T_a\int[\hat G^R_{cc}(\Omega+\omega)\hat T^+_a\hat G^{+-}_{ac}(\omega)-\hat G^R_{cc}(\omega)\hat T_a\hat G^{+-}_{ac}(\Omega$$
$$+\omega)]\frac{d\omega}{2\pi}=\hat T_a\int\hat G^R_a(\omega)\hat T^+_a\hat G^R_{cc}(\omega)[\hat G^{-+}_{cc}(\Omega+\omega)$$
$$-\hat G^{-+}_{cc}(\omega-\Omega)]\frac{d\omega}{2\pi}=N_a\langle|T_{ak_F}|^2\rangle[F(\varepsilon_c$$
$$+\Omega)\rho_c(\varepsilon_c+\Omega)]-[F(\varepsilon_c-\Omega)\rho_c(\varepsilon_c-\Omega)], \tag{16}$$

where the Keldysh function for the resonant level is found in analogy to Ref. 16:

$$G^{-+}_{cc}(\omega)=2\pi[1-F(\omega)]\rho_c(\omega),$$
$$F(\omega)=[\gamma_af(\omega)+\gamma_bf_b(\omega)]/\gamma, \tag{17}$$
$$\rho_c(\omega)=-\frac{1}{\pi}\,\mathrm{Im}\{G^R_{cc}(\omega)\}.$$

Here $f_a(\omega)$ and $f_b(\omega)$ are the Fermi distribution functions in the reservoirs (the substrate and the tip) with the chemical potentials μ_a and μ_b, respectively. If there is an applied voltage, we assume that it is included in the chemical potentials: $\mu_a-\mu_b=\mathrm{eV}$.

The second term in Eq. (12) can be brought into the form

$$\hat T_a\int[\hat G^R_{cc}(\Omega+\omega)\hat T^+_a\hat G^{+-}_{ac}(\omega)-\hat G^R_{cc}(\omega)\hat T^+_a\hat G^{+-}_{ac}(\Omega$$
$$-\omega)]\frac{d\omega}{2\pi}=\hat T_a\int[\hat G^{+-}_a(\omega)\hat G^R_{cc}(\omega)$$
$$+\hat G^A_a(\omega)\hat G^{+-}_{cc}(\omega)]\hat T^+_a[\hat G^A_{cc}(\omega+\Omega)-\hat G^{+-}_{cc}(\omega$$
$$-\Omega)]\frac{d\omega}{2\pi}=N_a\langle|T_{ak}|^2\rangle 2\,f_a(\varepsilon_c)\,\frac{\Omega}{\Omega^2+(2\gamma)^2}. \tag{18}$$

The functions G^R and G^A, which contain the polarization operators Π_{1b} and Π_{2b}, are calculated similarly. If the level of the extra electron ε_c in the particle is above the chemical potentials in the reservoirs when $V=0$ (which should be the case, as a rule, because of Coulomb repulsion), the terms with $f_{a,b}(\varepsilon_c)$ can be omitted. The final expression is simplified and is written as follows:

$$\langle II\rangle_\Omega = \text{const}\ J^2 N_a N_b |T_a|^2 |T_b|^2 \left(\frac{\Omega}{\Omega^2+(2\gamma)^2}\right)^2 \times \coth\left(\frac{\Omega}{2T}\right)\delta(\Omega-\Omega_{\mathrm{MQT}}). \quad (19)$$

The insignificant constants are included in the cofactor const. Recalling that $\gamma \propto |T_a|^2 N_a + |T_b|^2 N_b$ and that $\Omega_{\mathrm{MQT}} \ll \gamma$ [since $\gamma \approx 1$ meV and $\Omega_{\mathrm{MQT}} \approx 1$ kHz (Ref. 4)], we can bring Eq. (19) into the form

$$\langle II\rangle_\Omega \propto J^2 \left(\frac{\gamma\Omega}{\Omega^2+(2\gamma)^2}\right)^2 \coth\left(\frac{\Omega}{2T}\right)\delta(\Omega-\Omega_{\mathrm{MQT}}). \quad (20)$$

The appearance of a spike at Ω_{MQT} is not surprising, since the presence of an eigenmode in the system should be manifested in the fluctuation spectrum of quantities assigned to the combined substrate+particle+tip system. In a certain sense it is analogous to the appearance of a sharp feature (with weak damping) at the natural resonant frequency of an ordinary electrical oscillator circuit.

3. CONCLUSIONS

A simplified model of a magnetic particle with one spin on a site has been considered previously (see, for example, Ref. 8). We shall now show that this restriction is not of fundamental importance. To illustrate this, it is convenient to use the Heisenberg model with an anisotropic interaction. As was shown in Ref. 21, if there is strong anisotropy [$J_\perp \ll J_\parallel$, where $J_\perp$ and $J_\parallel$ are the exchange constants describing the exchange processes $J_\perp(S_{ix}S_{jx}+S_{iy}S_{jy})$ and $J_\parallel S_{iz}S_{jz}$], in this case the wave function of the ground state is

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\uparrow\downarrow\ldots\rangle + |\downarrow\uparrow\downarrow\uparrow\ldots\rangle).$$

The function corresponding to the first excited state is

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\uparrow\downarrow\ldots\rangle - |\downarrow\uparrow\downarrow\uparrow\ldots\rangle).$$

The magnitude of the splitting between the ground and the first excited state is exponentially small in the number of particles (N) and the ratio of the constants $J_\perp/J_\parallel$ and equals

$$\Delta \approx \omega^+ - \omega^- \propto J_\perp \exp(-N \ln(J_\parallel/J_\perp)).$$

When an electron tunnels from the substrate into the tip through the particle, an interaction with one of the spins in the particle $\hat{S}_i$ takes place. Special terms arise from the series expansion of the $\hat{S}$ matrix in (6) from terms of the form

$$\sum_i \left\langle T_{ia}T_{ia}^* T_{ib}T_{ib}^* \ldots \hat{c}_{iz}^+(\tau_3)\hat{c}_{iz}(\tau_3)\hat{S}_{iz}(\tau_3)\hat{S}_{iz}(\tau_4)\hat{c}_{iz}^+(\tau_4) \times \hat{c}_{iz}(\tau_4)\right\rangle \propto \sum_k \langle \Psi^n|\hat{S}_{iz}(\tau_3)\hat{S}_{iz}(\tau_4)|\Psi^n\rangle, \quad (21)$$

where T_{ia} and T_{ib} are the tunneling matrix elements that describe the passage of an electron from the substrate and the needle to the ith site in the magnetic particle, and $|\Psi^n\rangle$ denotes the eigenfunctions of the nth particular state of the particle. In view of the fact that[21]

$$\hat{S}_{iz}|\Psi^+\rangle = \pm S|\Psi^-\rangle, \quad \hat{S}_{iz}|\Psi^-\rangle = \pm S|\Psi^+\rangle, \quad (22)$$

the sign (plus or minus) in (22) is determined simply by the order of the labels of the spins in the state $|\uparrow\downarrow\uparrow\downarrow\ldots\rangle$. Substituting the expressions for the spin operators in the Heisenberg representation

$$\hat{S}_{iz}(\tau_{3,4}) = \exp(-i\hat{H}_{\mathrm{spin}}\tau_{3,4})\hat{S}_{iz}\exp(i\hat{H}_{\mathrm{spin}}\tau_{3,4}) \quad (23)$$

into (21) and recalling that $|\Psi^+\rangle$ and $|\Psi^-\rangle$ are the eigenstates of the Hamiltonian $\hat{H}_{\mathrm{spin}}$ of the particle, we have

$$\langle\Psi^+|\hat{S}_{iz}(\tau_3)\hat{S}_{iz}(\tau_4)|\Psi^+\rangle = \exp(i\Delta(\tau_3-\tau_4))|\langle\Psi^-|\hat{S}_{iz}|\Psi^+\rangle|^2. \quad (24)$$

The subsequent transition to the Fourier representation in (21) with consideration of Eq. (24) gives a peak in (21) at the magnetic macroscopic quantum tunneling frequency in the particle ($\Omega_{\mathrm{MQT}}=\Delta$). The contribution from each ith site with the spin S_i has its own weight, which depends on the squares of the tunneling matrix elements coupling the respective site in the particle with states in the substrate and the scanning tunneling microscope tip.

The contribution of macroscopic quantum tunneling to the spectrum of tunneling current fluctuations appears even in the absence of a spin–orbit interaction in the tip. The aforementioned tunneling current fluctuations, as well as the amplitude of the EPR signal in a scanning tunneling microscope, are proportional to the square of the spin–orbit interaction in the scanning tunneling microscope tip.[14,16] This is because the spin states $|\uparrow\rangle$ and $|\downarrow\rangle$ on a paramagnetic center in a constant magnetic field are eigenstates. The energy difference between these states is equal to the Zeeman energy (the Larmor frequency). When electrons tunnel through a paramagnetic center, spin-flip transitions of the electrons are required for the Larmor frequency to be manifested in the spectrum of tunneling current fluctuations. Because the tunneling current operator is diagonal with respect to the spin, ''coupling'' of the electron spin degrees of freedom can occur only to the extent of the spin–orbit interaction. In the case of macroscopic quantum tunneling in a magnetic particle, the states

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\uparrow\downarrow\ldots\rangle + |\downarrow\uparrow\downarrow\uparrow\ldots\rangle) \text{ and } |\Psi^-\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\uparrow\downarrow\ldots\rangle - |\downarrow\uparrow\downarrow\uparrow\ldots\rangle),$$

which differ by the magnitude of the macroscopic quantum tunneling energy, already contain components with different spin projections on a given site (the spin in the site does not have a definite projection). Therefore, a tunneling electron with a given spin projection ''finds'' the appropriate spin projection on the site. This means that a spin-flip transition of the electron and the presence of a spin–orbit interaction are not required to ''couple'' the spin degrees of freedom with the tunneling current.

Thus, the scanning tunneling microscope can be an effective tool for observing macroscopic quantum tunneling, especially because, as was previously shown, the sensitivity

of this microscope is sufficient even for detecting tunneling current fluctuations associated with the precession of an individual electron spin on a surface.[13]

In conclusion, we thank S. V. Iordanskiĭ and S. S. Nazin for fruitful discussions during the performance of this research. This work was supported by the Russian Fund for Fundamental Research (Project No. 94-02-04843), as well as the ''Surface Atomic Structures'' Program (Project No. 95-1.1).

[1] C. P. Bean and J. D. Livingston, J. Appl. Phys. **30**, 120 (1959).
[2] E. Schrödinger, Naturwissenschaften **23**, 884 (1935).
[3] E. M. Chudnovskiĭ, Zh. Éksp. Teor. Fiz. **77**, 2157 (1979) [Sov. Phys. JETP **50**, 1035 (1979)].
[4] D. D. Awschalom, D. P. DiVincenzo, and J. F. Smyth, Science **258**, 414 (1992); D. D. Awschalom and D. P. DiVincenzo, Phys. Today **43**, (1995).
[5] B. Barbara and E. M. Chudnovsky, Phys. Lett. A **145**, 205 (1990).
[6] E. M. Chudnovsky and L. Gunther, Phys. Rev. Lett. **60**, 661 (1990).
[7] P. C. E. Stamp, E. M. Chudnovsky, and B. Barbara, Int. J. Mod. Phys. B **6**, 1355 (1992).
[8] D. Loss, D. DiVincenzo, and G. Grinstein, Phys. Rev. Lett. **69**, 3232 (1992).
[9] D. D. Awschalom, J. F. Smyth, G. Grinstein, D. P. DiVincenzo, and D. Loss, Phys. Rev. Lett. **68**, 3092 (1992); D. D. Awschalom, M. A. McCord, and G. Grinstein, Phys. Rev. Lett. **65**, 783 (1990).
[10] B. Barbara, L. C. Sampaio, J. E. Wegrowe, B. A. Ratnam, A. Marchand, C. Paulsen, M. A. Novak, J. L. Tholence, M. Uehara, and D. Fruchart, J. Appl. Phys. **73**, 6703 (1993).
[11] J. Tejada, X. X. Zhang, and L. I. Ballsells, J. Appl. Phys. **73**, 6709 (1993).
[12] I. V. Shvets, R. Wiesendanger, D. Brügler, G. Tarrach, H.-J. Güntherodt, and J. M. D. Coey, J. Appl. Phys. **71**, 5496 (1992); R. Wiesendanger, H.-J. Güntherodt, G. Güntherodt, and R. Ruf, Phys. Rev. Lett. **65**, 247 (1990).
[13] Y. Manassen, R. J. Hamers, J. E. Demuth, and A. J. Castellano Jr., Phys. Rev. Lett. **62**, 2513 (1989).
[14] S. N. Molotkov, Surf. Sci. **264**, 235 (1992); S. N. Molotkov, JETP Lett. **59**, 190 (1994).
[15] A. W. McKinnon and M. E. Weland, in *Abstracts of STM '91, International Conference of Scanning Tunneling Microscopy*, Interlaken, 1991, p. 51.
[16] S. N. Molotkov, Surf. Sci. **302**, 235 (1994).
[17] S. N. Molotkov, JETP Lett. **55**, 173 (1992).
[18] M. W. Prins, R. Jansen, and H. van Kempen, Phys. Rev. B **53**, 8105 (1996).
[19] C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, J. Phys. C: Solid State Phys. **4**, 916(1971); **5**, 21 (1972).
[20] L. Y. Chen and C. S. Ting, Phys. Rev. B **43**, 2097 (1991).
[21] S. N. Molotkov and S. S. Nazin, JETP Lett. **63**, 718 (1996).

Translated by P. Shelnitz

Anomalous light absorption in molecular crystals

V. I. Tovstenko

Institute of Physics, Ukrainian National Academy of Sciences, 252650 Kiev, Ukraine Russia
(Submitted 13 November 1995; resubmitted 17 March 1997)
Zh. Éksp. Teor. Fiz. **112**, 1021–1040 (September 1997)

The method of successive approximations is used to obtain a solution for an infinite chain of nonlinear coupled equations for the Fourier transform of the retarded two-time Green's function in a system in which the electron vibrational interactions of the lattice photons and of the intramolecular photons are included in the terms of the Hamiltonian that are diagonal in the electron operators. A unique subsequence, which provides a finite contribution to the integral intensity of the absorption spectrum, is extracted from the infinite series and summed. The resulting formulas are used to classify the excited states of a crystal and interpret the cases of anomalous light absorption, when the absorption intensity becomes inversely proportional to the Franck–Condon overlap integral.
 [S1063-7761(97)01709-5]

1. INTRODUCTION

Although at present there exists an enormous body of theoretical and experimental data on the problem of exciton–phonon interaction in molecular crystals[1–4] (a detailed bibliography of the work in this area of research can be found in the cited monographs) and several laws explaining the peculiarities of the absorption and luminescence spectra have been established, some experimental data have found no satisfactory explanation within the framework of modern theoretical ideas. Let us discuss the topic more thoroughly.

As is known,[5,6] in the Franck–Condon approximation the intensity of an optical transition in a crystal from the ground electron state to an excited state is determined by what is known as the overlap integral $\langle \Psi^f | \Psi^0 \rangle$, where Ψ^0 and Ψ^f are the vibrational wave functions of the ground and excited electron states, respectively. In particular, in a phononless transition the overlap integral has a structure of the form $\exp(-u)$, where u is proportional to the product of the exciton–phonon coupling constant and the average number of phonons $\langle n \rangle$ (some researchers[7] call $\exp(-u)$ the Debye–Waller factor, by analogy with the probability of coherent (Bragg) scattering of x-rays and neutrons by a crystal).

Since the linear exciton–phonon coupling constant is proportional to the square root of the effective mass of the molecule (nucleus),[2,6] in isotope-substituted crystals at low temperatures the intensity of light absorption at the phononless resonance maximum must decrease with increasing isotope mass. If the temperature shift of the level is insignificant, the absorption intensity must behave in the same way with increasing temperature. At high temperatures the effect of the mass factor grows, so that the root dependence is transformed into linear dependence.

Some experimental data, however, point to an opposite dependence. For instance, for the deuteronaphthalene (D_8) crystal at liquid-helium temperatures the absorption coefficient at the maximum of the **a**-polarized phononless band is approximately 20% higher than the absorption coefficient for the naphthalene (H_8) crystal.[8] Robinette *et al.*[9] firmly established that within in a certain temperature interval (up to 10 K) and at the maximum of the **a**-band the absorption coefficient grows with temperature without a noticeable shift in the resonance on the frequency scale.

Since the exciton–phonon interaction renormalizes the exciton band by the factor $\exp(-u)$ (Ref. 4), in isotope-substituted crystals one should also observe a dependence of the Davydov splitting on the type of isotope: the splitting becomes smaller as isotope mass grows. In experiments, at least in aromatic crystals, no such dependence has been observed.[8,10,11]

There is also a group of experimental facts in which the observed anomalies are not directly related to the factor $\exp(-u)$. For instance, because of the weakness of the intermolecular interaction, the difference in the values of the polarization ratio for the components of a Davydov doublet in a crystal and in the oriented-gas model (without allowance for the interaction) must act as a small correction. For the naphthalene crystal, the polarization ratio exceeds that for a gas by a factor greater than 100.

Moreover, since in molecular crystals the parameters of the exciton bands are usually determined by the intramolecular multipole moments of transitions from the ground state to an excited state,[1] the value of the Davydov splitting in naphthalene does not agree with the oscillator strength of an intramolecular transition, i.e., it is anomalously large ($\simeq 160\ \mathrm{cm}^{-1}$) and comparable to a similar value for the anthracene crystal ($\simeq 220\ \mathrm{cm}^{-1}$) (Ref. 2; the oscillator strengths of intramolecular transitions in these crystals differ by several orders of magnitude[2]).

These examples and other data[12–19] suggest that the problem of the absorption bands in molecular crystals is in no way trivial and merits a more thorough study.

We believe that the discrepancy between theory and experiment arises because the calculations were not very accurate. For instance, some properties of the system, say, its nonlinearity in the problem with exciton–phonon interaction, are lost when the chain of the coupled equations for the retarded Green's functions is truncated at an early stage[2] or when terms in which the number of particles are not con-

served are discarded in model Hamiltonians,[4,20,21] to name some of the reasons.

Note that during recent years a number of new methods have appeared in the theory of the exciton–phonon interaction, methods that not only made it possible to take into account the polaron shift of the electron state but also provide a correct way of calculating the effect of lattice vibrations on the polaron band structure.[22–24] However, the papers we have just cited and other work devoted to this topic[25–27] yielded no material that would be useful in establishing a correspondence between the existing theory and the experimental facts listed at the beginning of this Introduction.

The present paper suggests a method to solve this problem. It is based on using the complete chain of coupled equations for the two-time retarded Green's function. With it we derive an expression for Green's functions in the form of a series in increasing powers of time-independent correlation functions. Partial splitting of the correlation functions for all terms in the series and summation of the resulting subsequences makes it possible to derive formulas that are in good agreement with the experimental data. New features, not mentioned previously by other researchers, have also been discovered.

2. THE CRYSTAL HAMILTONIAN. THE GREEN'S FUNCTION

In our studies we will find it convenient to use the mixed ($\mathbf{k}$–$\mathbf{n}$)-representation,[28] where $\mathbf{k}$ is the wave vector of a lattice phonon, and $\mathbf{n}$ is the radius vector of a molecule in the crystal. Generalizing the results of Refs. 1–4, 20 and 21, we can write the following expression for the Hamiltonian of a molecular crystal in which electron excitons interact both with intramolecular phonons and with crystal phonons, without allowance for electron configuration and the Duschinsky effect[29] in the harmonic approximation ($\hbar=1$):

$$H=H_R+\sum_{\mathbf{n}s}\ \omega^0_{\mathbf{n}s}a^\dagger_{\mathbf{n}s}a_{\mathbf{n}s}+\sum_{\mathbf{n}}\ B^\dagger_{\mathbf{n}}B_{\mathbf{n}}\Bigg(\varepsilon^0_{\mathbf{n}}+\sum_s\ \Delta\omega_{\mathbf{n}s}a^\dagger_{\mathbf{n}s}a_{\mathbf{n}s}\Bigg)+\sum_{\mathbf{nm}}{}'M^0_{\mathbf{nm}}V^\dagger_{\mathbf{m}}V_{\mathbf{n}}. \tag{1}$$

Here

$$H_R=\sum_{\mathbf{k}r}\ \omega_{\mathbf{k}r}b^\dagger_{\mathbf{k}r}b_{\mathbf{k}r} \tag{2}$$

is the part of the phonon Hamiltonian describing the motion of molecules as a whole; $\omega_{\mathbf{k}r}$ is the dispersion of the rth branch; $b^\dagger_{\mathbf{k}r}$ and $b_{\mathbf{k}r}$ are the lattice phonon creation and annihilation operators; $\Delta\omega_{\mathbf{n}s}=\omega^f_{\mathbf{n}s}-\omega^0_{\mathbf{n}s}$ is the difference in frequencies (the frequency defect) of the sth normal intramolecular vibration in the excited ($\omega^f_{\mathbf{n}s}$) and ground ($\omega^0_{\mathbf{n}s}$) electron states with allowance for crystalline corrections; $a^\dagger_{\mathbf{n}s}$ and $a_{\mathbf{n}s}$ are the intramolecular phonon creation and annihilation operators (the operators $b^\dagger_{\mathbf{k}r}, b_{\mathbf{k}r}$ and $a^\dagger_{\mathbf{n}s}, a_{\mathbf{n}s}$ commute since they act on different variables); $\varepsilon^0_{\mathbf{n}}$ is the renormalized electron state of a molecule in the crystal; the $M^0_{\mathbf{nm}}$ are the integrals of excitation exchange between the molecules $\mathbf{n}$ and $\mathbf{m}$ in the Franck–Condon approximation; the prime on the summation sign indicates that summation is only over sites $\mathbf{n}\neq\mathbf{m}$; and $B^\dagger_{\mathbf{n}}V^\dagger_{\mathbf{n}}=B^\dagger_{\mathbf{n}}P^\dagger_{\mathbf{n}}F^\dagger_{\mathbf{n}}$ and $B_{\mathbf{n}}V_{\mathbf{n}}=B_{\mathbf{n}}P_{\mathbf{n}}F_{\mathbf{n}}$ are the creation and annihilation operators for an electron excitation at site $\mathbf{n}$ accompanied by deformation of the intramolecular configuration of the nuclei and the lattice. The electron operators $B^\dagger_{\mathbf{n}}$ and $B_{\mathbf{n}}$ at low concentrations of excited molecules can be considered Bose operators:[1,2]

$$\begin{aligned}
&P_{\mathbf{n}}=\prod_s\ P_{\mathbf{n}s},\\
&P_{\mathbf{n}s}=\exp[-\alpha_{\mathbf{n}s}(a^\dagger_{\mathbf{n}s}-a_{\mathbf{n}s})]\exp[-\kappa_{\mathbf{n}s}(a^\dagger_{\mathbf{n}s}a^\dagger_{\mathbf{n}s}-a_{\mathbf{n}s}a_{\mathbf{n}s})]\\
&\quad=\exp[\alpha^2_{\mathbf{n}s}(1+2\beta_{\mathbf{n}s})/2]\exp[-\alpha_{\mathbf{n}s}(1+2\beta_{\mathbf{n}s})a^\dagger_{\mathbf{n}s}]\\
&\qquad\times\exp(-\beta_{\mathbf{n}s}a^\dagger_{\mathbf{n}s}a^\dagger_{\mathbf{n}s})\exp(-\gamma_{\mathbf{n}s}a^\dagger_{\mathbf{n}s}a_{\mathbf{n}s})\\
&\qquad\times\exp[\alpha_{\mathbf{n}s}\ \exp(-\gamma_{\mathbf{n}s})a_{\mathbf{n}s}]\exp(\beta_{\mathbf{n}s}a_{\mathbf{n}s}a_{\mathbf{n}s}),
\end{aligned}\tag{3}$$

where $\alpha_{\mathbf{n}s}$ is the dimensionless coupling constant representing the linear vibronic interaction in the crystal and $\kappa_{\mathbf{n}s}$ is the constant of quadratic vibronic coupling,

$$\kappa_{\mathbf{n}s}=\frac{1}{4}\ln\frac{\omega^f_{\mathbf{n}s}}{\omega^0_{\mathbf{n}s}},\quad \beta_{\mathbf{n}s}=\frac{1}{2}\tanh 2\kappa_{\mathbf{n}s}=\frac{1}{2}\frac{\omega^f_{\mathbf{n}s}-\omega^0_{\mathbf{n}s}}{\omega^f_{\mathbf{n}s}+\omega^0_{\mathbf{n}s}},$$

$$\exp(-\gamma_{\mathbf{n}s})=\sqrt{1-4\beta^2_{\mathbf{n}s}}. \tag{4}$$

In an ideal (translation-invariant) crystal the quantities $\alpha_{\mathbf{n}s}$, $\kappa_{\mathbf{n}s}$, and $\varepsilon^0_{\mathbf{n}}$ are independent of the molecule's number.

The operator $F_{\mathbf{n}}$ describes the deformation of the crystal lattice and has the form[3,28]

$$\begin{aligned}
F_{\mathbf{n}}&=\exp\Bigg[\sum_{\mathbf{k}r}\ (D^{\mathbf{k}r}_{\mathbf{n}}b_{\mathbf{k}r}-(D^{\mathbf{k}r}_{\mathbf{n}})^*b^\dagger_{\mathbf{k}r})\Bigg]\\
&=\prod_{\mathbf{k}r}\ \exp\Bigg(-\frac{|D_{\mathbf{k}r}|^2}{2}\Bigg)\exp[-(D^{\mathbf{k}r}_{\mathbf{n}})^*b^\dagger_{\mathbf{k}r}]\exp(D^{\mathbf{k}r}_{\mathbf{n}}b_{\mathbf{k}r}).
\end{aligned}\tag{5}$$

Here

$$D^{\mathbf{k}r}_{\mathbf{n}}=D_{\mathbf{k}r}\ \exp(i\mathbf{k}\cdot\mathbf{n}),$$

where $D_{\mathbf{k}r}$ is the Fourier component of the linear exciton–photon coupling constant determined by using the matrix elements of the potential energy of the interaction of a molecule and its surroundings.[2,3] We ignore the spatial dependence in the resonant terms since they are higher-order quantities in relation to $D_{\mathbf{k}r}$ (Ref. 30).

In the problem under consideration, the spectral properties of the crystal can be shown[4,21] to be described by the Fourier transforms of the following Green's functions:

$$G_{\mathbf{nm}}(t)=-i\,\theta(t)\langle[e^{iHt}B_{\mathbf{n}}V_{\mathbf{n}}e^{-iHt},\ B^\dagger_{\mathbf{m}}V^\dagger_{\mathbf{m}}]\rangle \tag{6}$$

allows for Franck–Condon transitions, and

$$\begin{aligned}
G^{HT}_{\mathbf{n}s,\mathbf{n};\mathbf{m},\mathbf{m}s'}(t)=-i\,\theta(t)\langle[e^{iHt}(a_{\mathbf{n}s}+a^\dagger_{\mathbf{n}s})B_{\mathbf{n}}V_{\mathbf{n}}e^{-iHt},\\
\times B^\dagger_{\mathbf{m}}V^\dagger_{\mathbf{m}}(a_{\mathbf{m}s'}+a^\dagger_{\mathbf{m}s'})]\rangle
\end{aligned}\tag{7}$$

allows for the Herzberg–Teller interaction.

The function (7) is the consequence of allowing for terms that are linear in the displacements from the equilib-

rium positions of the nuclei in the expansion of the matrix elements of the intramolecular transition current from the ground electron state to an excited state.

The Herzberg–Teller Green's function makes it possible to examine transitions to odd-parity sublevels in the energy spectrum involving inncompletely symmetric vibrations forbidden in the Franck–Condon approximation. In this case, according to the Jahn–Teller effect,[31] for nondegenerate electron states the symmetry of the molecule is conserved (the points of electron minima of the ground and excited states coincide), and the principal mechanism of interaction between the electron and vibrational motions (in contrast to completely symmetric vibrations, where the change of position of equilibrium plays the key role) is the phonon frequency shift. This effect is not small and amounts to several tens of cm^{-1} (see Ref. 4). Moreover, in some molecules of the aromatic series the intensity of a transition involving an incompletely symmetric vibration exceeds the intensity of a phononless transition by several orders of magnitude,[4,8] with the result that such transitions play an important role in the general picture of the absorption spectrum of a crystal.

Since the energy of intramolecular excitations is much higher than the temperatures at which experiments are usually conducted, the averaging in (6) and (7) for the electron subsystem and the subsystem of intramolecular phonons can be done over the ground state. For the subsystem of lattice phonons the averaging is done over the Gibbs canonical ensemble. Thus, the problem is reduced to calculating simplified Green's functions of the form

$$G_{\mathbf{mn}}(t)=-i\theta(t)\langle\exp(iH_R t)B_{\mathbf{n}}V_{\mathbf{n}}\exp(-iHt)B_{\mathbf{m}}^{\dagger}V_{\mathbf{m}}^{\dagger}\rangle, \tag{8}$$

$$G^{HT}_{\mathbf{ns},\mathbf{n};\mathbf{m},\mathbf{ms}'}(t)=-i\theta(t)\langle\exp(iH_R t)a_{\mathbf{ns}}B_{\mathbf{n}}V_{\mathbf{n}}\times\exp(-iHt)B_{\mathbf{m}}^{\dagger}V_{\mathbf{m}}^{\dagger}a_{\mathbf{ms}'}^{\dagger}\rangle. \tag{9}$$

The equations for the Green's functions (8) and (9) are solved by the method of successive approximations in Appendix A. Collecting the resulting expressions of the form (A8) and (A9) with subsequent iteration cycles, we have

$$\begin{aligned}G_{\mathbf{nm}}(E+i\gamma)=&\lim_{\gamma\to+0}\frac{1}{E-\varepsilon^0+i\gamma}\\&\times\sum_{\nu=0}^{\infty}\frac{\langle\tilde{K}_{\mathbf{n}}^{\nu}\rangle}{(E-\varepsilon^0+i\gamma)^{\nu}}\delta_{\mathbf{nm}}+\frac{M^0_{\mathbf{nm}}}{(E-\varepsilon^0+i\gamma)^2}\\&\times\sum_{\nu_1=0}^{\infty}\sum_{\nu_2=0}^{\infty}\frac{\langle\tilde{K}_{\mathbf{nm}}^{\nu_1\nu_2}\rangle}{(E-\varepsilon^0+i\gamma)^{\nu_1+\nu_2}}\\&+\sum_{\mathbf{n}_2}\frac{M^0_{\mathbf{nn}_2}M^0_{\mathbf{n}_2\mathbf{m}}}{(E-\varepsilon^0+i\gamma)^3}\\&\times\sum_{\nu_1=0}^{\infty}\sum_{\nu_2=0}^{\infty}\sum_{\nu_3=0}^{\infty}\frac{\langle\tilde{K}_{\mathbf{nn}_2\mathbf{m}}^{\nu_1\nu_2\nu_3}\rangle}{(E-\varepsilon^0+i\gamma)^{\nu_1+\nu_2+\nu_3}}\\&+\cdots+\sum_{\mathbf{n}_2}\cdots\sum_{\mathbf{n}_{j-1}}\frac{M^0_{\mathbf{nn}_2}\cdots M^0_{\mathbf{n}_{j-1}\mathbf{m}}}{(E-\varepsilon^0+i\gamma)^j}\\&\times\sum_{\nu_1=0}^{\infty}\cdots\sum_{\nu_j=0}^{\infty}\frac{\langle\tilde{K}_{\mathbf{nn}_2\cdots\mathbf{n}_{j-1}\mathbf{m}}^{\nu_1\nu_2\cdots\nu_{j-1}\nu_j}\rangle}{(E-\varepsilon^0+i\gamma)^{\nu_1+\cdots+\nu_j}}+\cdots.\end{aligned} \tag{10}$$

The expressions for the correlation functions $\langle\tilde{K}\rangle$ are given by formulas (A13)–(A15), (A19), and (A22) in Appendix A.

If in the correlator (A14) at $q=1$ $(\mathbf{n}_1=\mathbf{n})$ the operator $P_{\mathbf{n}_1}$ is replaced by $a_{\mathbf{ns}}P_{\mathbf{n}}$ and at $q=j$ $(\mathbf{n}_j=\mathbf{m})$ $P^{\dagger}_{\mathbf{n}_j}$ is replaced by $P^{\dagger}_{\mathbf{m}}a^{\dagger}_{\mathbf{ms}'}$, the series (10) is the solution for the Herzberg–Teller Green's function (9), too.

As (10) shows, the Fourier transform of the Green's function is an extremely complicated function of frequency, temperature, and spatial coordinates, so that at present there is no way in which the series can be fully summed. To analyze the series we introduce the following simplifications. According to (A27), each correlation function $\langle\tilde{K}\rangle$ contains a term that is independent of the spatial coordinates and, in accordance with the order of this function, can be represented as a power factor of the first term in (10). The terms specified in this way form a subsequence of the geometric progression type and can easily be summed.

If we note also that for a crystal with two molecules in the unit cell,

$$M^0_{\mathbf{nm}}\to M^0_{\mathbf{n}\sigma,\mathbf{m}\lambda}=\frac{1}{N\kappa}\sum_{\mathbf{Q}}\exp\{i\mathbf{Q}\cdot(\mathbf{n}-\mathbf{m})\}\times\sum_{\mu}u^{\mu}_{\sigma}(\mathbf{Q})[u^{\mu}_{\lambda}(\mathbf{Q})]^{*}\varepsilon_{\mathbf{Q}\mu}, \tag{11}$$

where u^{μ}_{σ} is the eigenfunction of the operator

$$M^0_{\sigma\lambda}=\sum_{\mathbf{n}}{}'\exp\{i\mathbf{Q}\cdot(\mathbf{n}-\mathbf{m})\}M^0_{\mathbf{n}\sigma,\mathbf{m}\lambda}, \tag{12}$$

κ is the number of molecules per unit cell, and μ is the label that numbers the exciton bands, summation over the spatial coordinates leads to the following expression for the Franck–Condon Green's function:

$$G_{\mathbf{n}\sigma,\mathbf{m}\lambda}(E)=\frac{1}{N\kappa}\sum_{\mathbf{Q}\mu}u^{\mu}_{\sigma}(\mathbf{Q})[u^{\mu}_{\lambda}(\mathbf{Q})]^{*}\times\exp\{i\mathbf{Q}\cdot(\mathbf{n}-\mathbf{m})\}G_{\mathbf{Q}\mu}(E). \tag{13}$$

Here

$$G_{\mathbf{Q}\mu}(E)=\frac{S_E}{1-\varepsilon_{\mathbf{Q}\mu}S_E}, \tag{14}$$

where

$$\begin{aligned}S_E=&\sum_{\nu=0}^{\infty}\frac{\langle\tilde{K}_{\mathbf{n}}^{\nu}\rangle}{(E-\varepsilon^0+i\gamma)^{\nu+1}}\\=&\prod_s\sum_{l_s}f_{l_s}F^2_{l_s}\prod_{\mathbf{kr}}\exp[-|D_{\mathbf{kr}}|^2(2\langle n_{\mathbf{kr}}\rangle+1)]\\&\times\sum_{\mu_{\mathbf{kr}}=0}^{\pm\infty}\frac{L_{\mu_{\mathbf{kr}}}}{E-\varepsilon^0-\cdots-\omega^f_s l_s-\cdots-\omega_{\mathbf{kr}}\mu_{\mathbf{kr}}-\cdots+i\gamma},\end{aligned} \tag{15}$$

$$L_{\mu_{\mathbf{k}r}}=\left(\frac{\langle n_{\mathbf{k}r}\rangle+1}{\langle n_{\mathbf{k}r}\rangle}\right)^{\mu_{\mathbf{k}r}/2} I_{\mu_{\mathbf{k}r}}(2|D_{\mathbf{k}r}|^2\sqrt{\langle n_{\mathbf{k}r}\rangle(\langle n_{\mathbf{k}r}\rangle+1)}), \tag{16}$$

$I_{\mu_{\mathbf{k}r}}$ is the modified Bessel function of order $\mu_{\mathbf{k}r}$,

$$f_{l_s}=\exp[-\alpha_s^2(1+2\beta_s)-\gamma_s]\,\frac{(\alpha_s^2\,\exp(-\gamma_s))^{l_s}}{l_s!}, \tag{17}$$

$$F_{l_s}=\sum_{\mu=0}^{E(l_s)}\left(\frac{\beta_s\,\exp(2\gamma_s)}{\alpha_s^2}\right)^{\mu}\frac{l_s!}{\mu!(l_s-2\mu)!}, \tag{18}$$

and $E(l_s)$ is the integral part of the number l_s.

If the sth intramolecular vibration is not completely symmetric and the crystalline correction to α_s is zero, i.e., $\alpha_s=0$ as a whole, we have

$$f_{l_s}F_{l_s}^2=\exp(-\gamma_s)\,\frac{\beta_s^{2l_s}}{l_s!l_s!}(2l_s)!, \tag{19}$$

and in the denominator of (15) $\omega_s^f l_s$ must be replaced by $\omega_s^f 2l_s$.

Now let us examine the Green's function (9), which describes Herzberg–Teller states. In this case, in (15) and (18) the expression for F_{l_s} becomes

$$\widetilde{F}_{l_s}=\sum_{p=0}^{E(l_s)}\left(\frac{\beta_s\,\exp(2\gamma_s)}{\alpha_s^2}\right)^{p}\frac{l_s!}{p!(l_s-2p)!}\times\left[\frac{l_s-2p}{\alpha_s}-\alpha_s(1+2\beta_s)\right] \tag{20}$$

and the sum S_E becomes

$$S_E^{ss'}=\delta_{ss'}\sum_{l_s}\,f_{l_s}\widetilde{F}_{l_s}^2 S_{l_s}, \tag{21}$$

where

$$S_{l_s}=\prod_{\chi\neq s}\sum_{l_\chi}\,f_{l_\chi}F_{l_\chi}^2\prod_{\mathbf{k}r}\times\exp[-|D_{\mathbf{k}r}|^2(2\langle n_{\mathbf{k}r}\rangle+1)]\times\sum_{\mu_{\mathbf{k}r}}^{\pm\infty}\frac{L_{\mu_{\mathbf{k}r}}}{E-\varepsilon^0-\omega_s^f l_s-\cdots-\omega_\chi^f l_\chi-\cdots-\omega_{\mathbf{k}r}\mu_{\mathbf{k}r}-\cdots+i\gamma}. \tag{22}$$

To identify the terms that provide a finite contribution to the integral intensity, in (21) we introduced the function $\delta_{ss'}$, since for $s\neq s'$ we have

$$\int\ \mathrm{Im}\ S_E^{ss'}\,dE=0.$$

What is summed with the first term in (10) is also a subsequence of a geometric progression, but constructed from terms that are diagonal in the index l_s. Thus,

$$G^{HT}_{\mathbf{n}\lambda,\mathbf{n}\lambda s;\mathbf{m}\sigma,\mathbf{m}\sigma\, s'}(E)=\frac{\delta_{ss'}}{N\kappa}\sum_{\mathbf{Q}\mu}\,u_\lambda^\mu(u_\sigma^\mu)^*\times\exp\{i\mathbf{Q}\cdot(\mathbf{n}-\mathbf{m})\}g^s_{\mathbf{Q}\mu}, \tag{23}$$

$$g^s_{\mathbf{Q}\mu}=\sum_{l_s}\,\frac{f_{l_s}\widetilde{F}_{l_s}^2 S_{l_s}}{1-\varepsilon_{\mathbf{Q}\mu}f_{l_s}F_{l_s}^2 S_{l_s}}. \tag{24}$$

If the sth intramolecular vibration satisfies $\alpha_s=0$, the coefficients $f_{l_s}\widetilde{F}_{l_s}^2$ vanish for all values $l_s\neq 2p_s+1$, and the $f_{l_s}F_{l_s}^2$ are finite only at $l_s=2p_s$ ($f_{l_s}F_{l_s}\widetilde{F}_{l_s}=0$ for all integral values of l_s). Then

$$f_{l_s}\widetilde{F}_{l_s}^2=\widetilde{f}_{l_s}=\exp(-3\gamma_s)\,\frac{\beta_s^{2l_s}}{l_s!l_s!}(2l_s+1)!, \tag{25}$$

$$G^{HT}_{\mathbf{n}\lambda,\mathbf{n}\lambda s;\mathbf{m}\sigma,\mathbf{m}\sigma\, s'}(E)=\delta_{ss'}\delta_{\mathbf{nm}}\delta_{\lambda\sigma}g_s, \tag{26}$$

$$g_s=\sum_{l_s}\,\widetilde{f}_{l_s}S_{2l_s+1}, \tag{27}$$

and only the first term in (10) contributes to the integral absorption intensity.

Thus, if an incompletely symmetric phonon participates in an electron vibrational transition, the Herzberg–Teller excited states of the crystal are localized. As S_{2l_s+1} implies, the shape of the band in this case has the following structure: the resonant part of absorption is determined by the intramolecular transition frequency, while the broadband part is determined by the spectral density of crystal vibrations. The so-called one-particle and dissociated (two-particle) states[4,20] are not considered here.

As Eqs. (14), (24), and (26) show, the Green's functions obtained from the subsequences summed with the first term of the series (10) have resonant denominators only of the first order. Hence each integral of their imaginary parts (the dissipative part of the dielectric constant) over the entire frequency range has a finite (nonzero) value. The other terms in (10) have higher-order resonance denominators, so that the subsequences obtained from these terms have the same property. These additional terms contribute nothing to the integral intensity, but both their absorption ($-\mathrm{Im}\,G'>0$) and emission ($-\mathrm{Im}\,G'<0$) parts slightly distort some fragments of the spectrum. To estimate these distortions, in Appendix B we extract and sum one such subsequence, based on the third term in (10) (we allowed only for one branch of the vibronic correlator (A13)). Comparison of (15) with (A32) and (A35) shows that the relative value of the correction is of order $\alpha^2\exp(-2\alpha^2)$. As α^2 varies continuously from zero to infinity, the product $\alpha^2\exp(-2\alpha^2)$ reaches its maximum at $\alpha^2=0.5$ and is approximately 0.2.

3. THE ENERGY SPECTRUM OF A CRYSTAL AND COMPARISON WITH EXPERIMENTAL DATA

We can easily show that the shape of the absorption band to first order in the external electromagnetic field ($\mathbf{q}_0$ is the wave vector of the light) is determined, with allowance for Eqs. (14), (24), and (27), is determined by the following expressions;

$$-\mathrm{Im}\,G_{\mathbf{q}_0\mu}(E)=-\frac{\mathrm{Im}\,S_E}{(1-\varepsilon_{\mathbf{q}_0\mu}\,\mathrm{Re}\,S_E)^2+(\varepsilon_{\mathbf{q}_0\mu}\,\mathrm{Im}\,S_E)^2}, \tag{28}$$

for Franck–Condon transitions,

$$-\operatorname{Im} g^{s}_{\mathbf{q}_0\mu} = -\sum_{l_s} \frac{f_{l_s}\widetilde{F}^2_{l_s} \operatorname{Im} S_{l_s}}{(1-\varepsilon_{\mathbf{q}_0\mu} f_{l_s} F^2_{l_s} \operatorname{Re} S_{l_s})^2 + (\varepsilon_{\mathbf{q}_0\mu} f_{l_s} F^2_{l_s} \operatorname{Im} S_{l_s})^2} \tag{29}$$

for Herzberg–Teller states involving the sth completely symmetric vibration, and

$$-\operatorname{Im} g_s = -\sum_{l_s} \widetilde{f}_{l_s} \operatorname{Im} S_{2l_s+1} \tag{30}$$

for Herzberg–Teller states involving the sth incompletely symmetric vibration.

According to (28) and (29), the energy levels of the normalized absorption spectrum can be found by solving the following equations:

$$1-\varepsilon_{\mathbf{q}_0\mu} \operatorname{Re} S_E = 0, \tag{31}$$

$$1-\varepsilon_{\mathbf{q}_0\mu} f_{l_s} F^2_{l_s} \operatorname{Re} S_{l_s} = 0. \tag{32}$$

The spectrum of Herzberg–Teller states involving an incompletely symmetric phonon is determined by the real part of the denominator of the function S_{2l_s+1} in (30).

In their mathematical structure, Eqs. (31) and (32) are similar to the Lifshitz equation,[32] which was first derived in the problem of degenerate regular perturbations of the discrete and quasicontinuous spectra. The main features of this equation have been thoroughly studied for different systems by many researchers (see, e.g., the monographs cited in Refs. 4, 33, and 34). If in (28) and (29) we formally identify the band component of the electron excitation, $\varepsilon_{\mathbf{q}_0\mu}$, with the potential that perturbs the spectrum of function S, we can use the method developed in Refs. 4 and 32–34 to study Eqs. (31) and (32).

To this end we must first simplify (15): in the part involving lattice photons we ignore the dependence on the wave vector of the optical branches, since the dispersion of these branches is low; for the acoustic branches we keep the terms linear in the exciton–phonon coupling constant; and for the intramolecular vibrations we put $\Delta\omega = 0$. Then Eq. (15) becomes

$$S_E \approx e^{-u} \prod_s \sum_{\nu_s} \frac{\alpha_s^{2\nu_s}}{\nu_s!} \prod_p \sum_{\mu_p=0}^{\pm\infty} L_{\mu_p}\left[\frac{1}{\Omega + i\gamma} + \sum_{\mathbf{k}r} \left|D_{\mathbf{k}r}\right|^2 \left(\frac{1+\langle n_{\mathbf{k}r}\rangle}{\Omega-\omega_{\mathbf{k}r}+i\gamma} + \frac{\langle n_{\mathbf{k}r}\rangle}{\Omega+\omega_{\mathbf{k}r}+i\gamma}\right)\right], \tag{33}$$

where

$$\Omega = E - \varepsilon^0 - \cdots - \omega_s^0 \nu_s - \cdots - \omega_p \mu_p - \cdots, \tag{34}$$

$$u = \sum_s \alpha_s^2 + \sum_p |D_p|^2 (1+2\langle n_p\rangle) + \sum_{\mathbf{k}r} |D_{\mathbf{k}r}|^2 (1+2\langle n_{\mathbf{k}r}\rangle). \tag{35}$$

Here the label p numbers the optical branches of the crystal lattice.

Let us inspect the second and third terms in the square brackets in (33). According to Davydov,[2]

$$D_{\mathbf{k}r} = \frac{1}{\hbar\omega_{\mathbf{k}r}} \sum_j e^j_{\mathbf{k}r} \left(\frac{\hbar}{2M_j\omega_{\mathbf{k}r}}\right)^{1/2} \times \sum_{\mathbf{m}} \left(\frac{\partial}{\partial u^j_0} + \exp(i\mathbf{k}\cdot\mathbf{m}) \frac{\partial}{\partial u^j_{\mathbf{m}}} D_{0\mathbf{m}}\right), \tag{36}$$

where the $e^j_{\mathbf{k}r}$ are the components of the unit polarization vectors of the rth branch with the wave vector $\mathbf{k}$, the M_j are the mass factors corresponding to the translational degrees of freedom of the molecule, $u^j_{\mathbf{m}}$ is the displacement of the molecule's center of gravity from the equilibrium position, and the matrix elements $D_{0\mathbf{m}}$ describe the change in the potential energy of the interaction of the molecule and its surroundings when the crystal goes into an excited state.

For a simple cubic lattice with the nearest-neighbors interaction at the low-frequency edge of the continuous spectrum (in an electron vibrational transition, according to Eq. (34), this edge is $\varepsilon = \varepsilon^0 + \cdots + \omega_s^0 \nu_s + \cdots + \omega_p \mu_p + \cdots$), the contribution of the second term to the density of states near the lower edge of the acoustic branch has the form ($\Omega > 0$)

$$g(E) \approx \text{const} \times \Omega^3 \frac{\exp(\Omega/kT)}{\exp(\Omega/kT)-1}. \tag{37}$$

At the high-frequency edge of the spectrum, where $\omega_m - \Omega \ll \omega_m$, , with ω_m the cutoff frequency of the acoustic branch, we have

$$g(E) \approx \text{const} \times \Omega^2 \sqrt{\omega_m^2 - \Omega^2}\, \frac{\exp(\Omega/kT)}{\exp(\Omega/kT)-1}. \tag{38}$$

Thus, in an electron vibrational transition, the density of phonon states at the edges of the continuous-spectrum band disappears according to different laws and differs from the density of ''purely'' acoustic vibrations of the crystal, with the latter density having a root singularity.[33]

This discrepancy is caused by effect that the modulating factors in the integrand depending on the wave vector, i.e., the constant $|D_{\mathbf{k}r}|^2$ and the average number of phonons $\langle n_{\mathbf{k}r}\rangle$, have on the spectral density and by the presence in the denominator of the Green's function of a term that is linear rather than quadratic in phonon dispersion.

In the frequency range $\Omega < 0$, the third term on the right-hand side of Eq. (33) contributes to the density of vibrations, with the corresponding formulas differing from (37) and (38) only by the absence of a temperature exponential in the numerator.

Since the density of the phonon states at the edges of the acoustic band vanishes, the real part of the terms under consideration decreases as $1/|\Omega|$ when $|\Omega| \gg \omega_m$. As we get closer to the band edges, the real part increases in absolute value, reaches a maximum, then decreases, and in the center of gravity of the distribution passes through zero and changes sign.[34] The size of the maximum near the band edges for the second term is of order

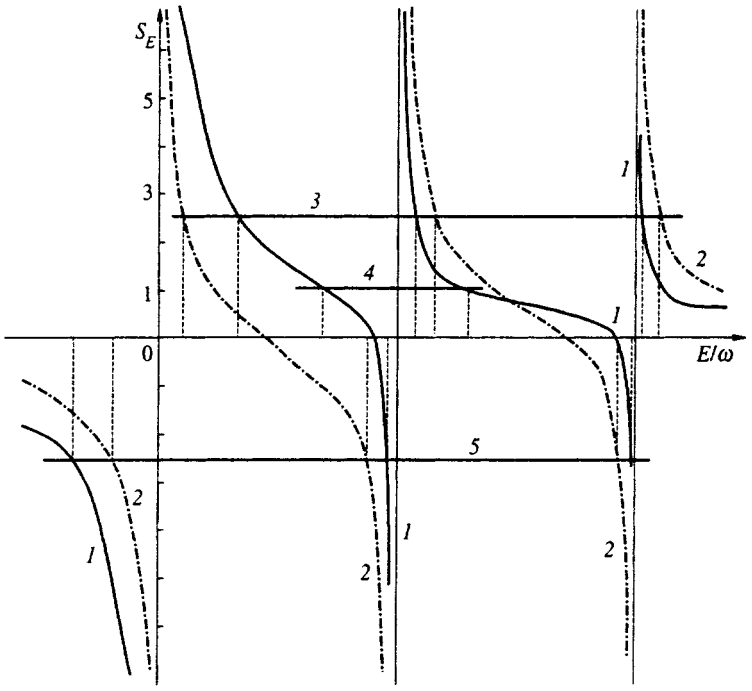

FIG. 1. Graphical solution of Eq. (31), $-S_E\varepsilon_{\mathbf{q}_0}/\omega=1$, in dimensionless units, where $S_E=\exp(-\alpha^2)\Sigma_\nu(\alpha^{2\nu}/\nu!)\times[(E-\varepsilon^0)/\omega-\nu]^{-1}$; curves *1* correspond to $\alpha^2=0.1$, curves *2* to $\alpha^2=1.0$, straight line *3* to $\omega/\varepsilon_{\mathbf{q}_0}=2.5$, straight line *4* to $\omega/\varepsilon_{\mathbf{q}_0}=1.0$, and straight line *5* to $\omega/\varepsilon_{\mathbf{p}_0}=-2.5$.

$$\mathrm{Re}\ S=C_2\approx\left|\frac{1}{(2\pi)^3}\int\frac{|D_{\mathbf{k}r}|^2(1+\langle n_{\mathbf{k}r}\rangle)}{\hbar\omega_{\mathbf{k}r}}d\mathbf{k}\right|,$$

and for the third, of order

$$\mathrm{Re}\ S=C_3\approx\left|\frac{1}{(2\pi)^3}\int\frac{|D_{\mathbf{k}r}|^2\langle n_{\mathbf{k}r}\rangle}{\hbar\omega_{\mathbf{k}r}}d\mathbf{k}\right|.$$

Since as we approach any of the discrete frequencies where Ω vanishes the real part of the first term changes sign and diverges and the second and third terms are finite, for $|1/\varepsilon_{\mathbf{q}_0}|>C_2$, C_3 the number of roots of Eq. (31) coincides with the complete set of discrete vibrations of the system. The second and third terms in this case act as corrections to the position of the level.

Now let us discuss in detail the distribution of the discrete roots in the zeroth approximation. To this end in (33) we isolate the part corresponding, say, to one of the intramolecular vibration branches. The qualitative picture describing the situation is depicted in Fig. 1. The first 20 terms in the sum

$$\mathrm{Re}\ S\approx\exp(-\alpha^2)\sum_\nu\frac{\alpha^{2\nu}}{\nu!}\frac{1}{E-\varepsilon^0-\omega\nu}\tag{39}$$

were used in the calculations, and point $E=\varepsilon^0$ was selected as the origin on the horizontal axis.

Figure 1 shows that, in contrast to an isolated molecule, in a crystal the vibrational series are not equidistant, with the greatest shift at the beginning of the spectrum. If $\varepsilon_{\mathbf{q}_0\mu}<0$, the energies of the vibrational sublevels are higher than the frequency of the initial phonon, and if $\varepsilon_{\mathbf{q}_0\mu}>0$, they are lower. For instance, at $\alpha^2=0.1$ (curves *1*) and $\omega/\varepsilon_{\mathbf{q}_0\mu}=1$ (straight line *4*), the contribution of the first electron-vibrational transition to the phonon frequency amounts to 0.6ω. As we move toward the violet part of spectrum, an increase in the occupation numbers leads to a reduction in the distortions of phonon frequencies, which is accompanied by a decrease in the Davydov splitting for polarization transitions. This effect of ''stretching'' or ''compressing'' the vibration frequency and the related nonequidistance of the series in the harmonic approximation are consequences of the nonlinearity of the equations for the Green's functions.

In the frequency range where the real part of the first term on the right-hand side of Eq. (33) is small (see Fig. 1), a situation may arise in which the acoustic branches provide the main contribution to Re S (usually ω_m does not exceed the lowest frequency of intramolecular or optical vibrations). If in this case $1/\varepsilon_{\mathbf{q}_0\mu}$ is smaller in absolute value than the constants C_2 and C_3 but has the same sign, Eqs. (31) and (32) may have roots in addition to the discrete frequencies (generally speaking, the solutions of Eqs. (31) and (32) do not coincide). The smaller the value of $-\mathrm{Im}\ S$ in this range (say, at the edges of the acoustic branch), the more pronounced are the corresponding resonances in the absorption spectrum.

Thus, our qualitative analysis suggests that each discrete sublevel that represents a transition involving intramolecular or optical phonons (with their dispersion ignored), including a purely excitonic transition, has accompanying broadband absorption whose extent is determined by the edges of the acoustic band. Moreover, within this band there may be additional resonances related to what is known as quasistationary states[33] (formally all states related to the solutions of Eqs. (31) and (32) are quasistationary because if we allow for multiphoton transitions in the acoustic branches of the spectrum at $T\neq0$, the spectral density is finite in the entire energy range).

Using the above reasoning, we can now unambiguously interpret the situation at the beginning of the absorption

spectrum of naphthalene (see the Introduction). Indeed, a narrow **a**-polarized resonance is an isolated solution of Eq. (31) in the region where $E<\varepsilon^0$ ($\varepsilon_{\mathbf{q}_0\mu}<0$). In **b**-polarization, $\varepsilon_{\mathbf{q}_0\mu}$ is positive, and the lowest solution of the equation lands in the acoustic band, so that it is not of resonance origin. On the other hand, the extent of broadband absorption and the positions of the maxima in both polarizations (they coincide, as shown in Refs. 15 and 17) are in good agreement with the data of Ref. 35 on the density of phonon states in the naphthalene crystal, so that the band, we believe, is primarily vibrational in origin.

Since the experimental data on oscillator strengths are based on measurements of the area under the vibrational curve, which is obviously an unsubstantiated method for extracting such data, the contradiction between the low oscillator strength in an isolated molecule and the anomalously high value in a crystal is resolved. Moreover, since in our interpretation the maximum in **b**-polarization is not a component of a Davydov doublet of a phononless transition, the contradiction related to the anomalously high value of the Davydov splitting is also resolved. Extremely accurate measurements are probably needed, to find a true **b**-component or its traces.

For a high oscillator strength, when the radiative width of an intramolecular transition becomes comparable to the width of the acoustic-phonon band and the frequencies of optical vibrations, identifying the components of a Davydov doublet is not so trivial. For instance, in the anthracene crystal, frequencies of six branches of optical lattice vibrations from 30 to 100 cm^{-1} fall into the electron transition band,[36] so that the resulting maximum in each component is formed by a superimposition of phonon wings overlapping in a complicated manner and the electron transition band proper.

The formulas derived in this paper also make it possible to explain the experimental facts, described in the Introduction, that concern the anomalous temperature dependence and the anomalies in isotope-substituted crystals. For instance, if $\varepsilon_{\mathbf{q}_0\mu}<0$, the solution to Eq. (31) related to an electron state lies in the frequency range $E<\varepsilon^0$, where the spectral phonon density at liquid-helium temperatures is formed primarily in quantum transitions from the vibrational sublevels of acoustic lattice vibrations to an electron excitation level and is therefore low. Since at low temperatures the principal contribution to the temperature dependence of Re S is provided by the terms linear in $\langle n\rangle$, which according to (33) have opposite signs, the position of the electron level does not change within a narrow temperature range. In this case the absorption intensity at the maximum of the resonance band is

$$-\operatorname{Im} G_{\mathbf{q}_0\mu}\propto -\frac{1}{(\varepsilon_{\mathbf{q}_0\mu})^2\operatorname{Im} S_E}\propto e^u, \tag{40}$$

so that the intensity increases with T.

Clearly, formula (40) also describes the increase in absorption intensity with isotope mass.

Let us take Herzberg–Teller states involving a not completely symmetric intramolecular phonon. In view of the localized nature of the excitation, the absorption intensity at the maximum of an isolated resonance, according to (30), always decreases with increasing temperature and with increasing isotope mass. This pattern is corroborated by the results of Sheka and Terenetskaya,[8] who studied the M-band of the naphthalene crystal. Since here the broadband part is the spectral manifestation of the density of phonon states, studying its structure may serve as a good addition to the method that uses neutrons,[35]

As for the experimentally observed fact that the Davydov splitting is independent of isotope mass,[8,10,12] we believe that comparison of theory and experiment at this stage is premature, since in this paper we disagree with the common interpretation of the bands.

An interesting experimental possibility of determining the state of a molecule in the crystal (the value of ε^0) follows from (15) and (28). These formulas show that at low temperatures the spectral density of phonons at point $E=\varepsilon^0$ is extremely low and that its temperature dependence is much weaker than in the other parts of the spectrum. Such a point can easily be seen in the figure for different temperature curves of the intensity of absorption in the naphthalene crystal in Ref. 17.

4. CONCLUSION

By using the method of two-time retarded Green's functions we have examined the effect of phonon excitations (both lattice and intramolecular) on the optical spectra of molecular crystal in the exciton absorption region. Specifically, we have studied the case where the electron vibrational interaction is accounted for by the terms in the Hamiltonian that are diagonal in the electron operators.

We have used the method of successive approximations to solve the chain of coupled nonlinear equations for the Fourier transform of the Green's function. The result is an infinite series in increasing order of phonon correlation functions. By partially splitting each correlation function we isolated and summed a unique subsequence that provides a finite contribution to the integral intensity of the absorption spectrum. We show that the effect of other subsequences, which contribute nothing to the integral intensity, reduces to a distortion of the spectrum by a value of about 20%.

Analysis of the results makes it possible to classify the excited states of the crystal as follows: all states except those formed by a Herzberg–Teller intramolecular transition involving an incompletely symmetric phonon are one-particle, and the latter are localized (one-particle in the sense that an electron excitation and an equilibrium deformation move through the crystal in a self-consistent manner as a whole, with a single wave vector).

The formulas derived here have made it possible to explain some anomalies in the absorption spectra of molecular crystals. In particular, for an excitonic resonance we have found that if it is shifted toward long waves from the molecular state, at low temperatures in a certain interval the intensity at the maximum of the band can be proportional to the reciprocal values of the overlap integral, i.e., must grow with temperature. In the same case the intensity strongly depends on the mass of the molecule in isotope-substituted

crystals: the intensity at the maximum of a resonance grows with isotope mass.

A new interpretation has been given for the lower part of the spectrum of the naphthalene crystal: **b**-polarized broadband absorption belongs not to a purely electron transition but is related to the density of long-wave phonon states of the crystal. Such an approach resolves the contradiction between the low oscillator strength of the transition in an isolated molecule and the incomparably higher intensity of absorption in the crystal, since in this case the common method of measuring the oscillator strength of a phononless transition by the area under the vibrational curve is unreliable. For the same reason the contradiction caused by the anomalously high value of the Davydov splitting is also resolved.

An experimental method is proposed for determining an important reference microparameter of a system, the molecular state in a crystal.

APPENDIX A: DERIVATION AND SOLUTION OF THE EQUATIONS FOR THE GREEN'S FUNCTIONS

We start by constructing an equation for the function (8). Differentiating with respect to time, we find that

$$i\frac{\partial}{\partial t}G_{\mathbf{nm}}(t)=\delta(t)\langle V_{\mathbf{n}}V_{\mathbf{m}}^{\dagger}\rangle\delta_{\mathbf{nm}}+\varepsilon^{0}G_{\mathbf{nm}}(t)-i\theta(t)\times\langle\exp(iH_{R}t)[V_{\mathbf{n}}H_{\mathbf{n}}]B_{\mathbf{n}}\times\exp(-iHt)B_{\mathbf{m}}^{\dagger}V_{\mathbf{m}}^{\dagger}\rangle-i\theta(t)\sum_{\mathbf{l}}{}'M_{\mathbf{nl}}^{0}\times\langle\exp(iH_{R}t)V_{\mathbf{n}}V_{\mathbf{n}}^{\dagger}B_{\mathbf{l}}V_{\mathbf{l}}\times\exp(-iHt)B_{\mathbf{m}}^{\dagger}V_{\mathbf{m}}^{\dagger}\rangle. \tag{A1}$$

Here

$$H_{\mathbf{n}}=H_{R}+H_{\mathbf{n}}^{v}, \tag{A2}$$

$$H_{\mathbf{n}}^{v}=\sum_{l_s}\omega_{s}^{0}a_{l_s}^{\dagger}a_{l_s}+\sum_{s}\Delta\omega_{s}^{0}a_{\mathbf{n}s}^{\dagger}a_{\mathbf{n}s}. \tag{A3}$$

Allowing for different averaging rules for the two phonon subsystems, in the third term on the right-hand side of Eq. (A1) we introduced, for symmetry considerations, an additional term $-H_{\mathbf{n}}^{v}V_{\mathbf{n}}$, which has no effect on the final result.

Differentiating this term with respect to time and then continuing the process indefinitely, we obtain a chain of nonlinear equations in which the νth equation has the following form:

$$i\frac{\partial}{\partial t}G_{\mathbf{nm}}^{\nu}(t)=\delta(t)\langle\widetilde{K}_{\mathbf{n}}^{\nu}\rangle\delta_{\mathbf{nm}}+\varepsilon^{0}G_{\mathbf{nm}}^{\nu}(t)+G_{\mathbf{nm}}^{\nu+1}(t)+\sum_{\mathbf{l}}{}'M_{\mathbf{nl}}^{0}G_{\mathbf{n,lm}}^{\nu}(t). \tag{A4}$$

Here

$$\widetilde{K}_{\mathbf{n}}^{\nu}=K_{\mathbf{n}}^{\nu}V_{\mathbf{n}}^{\dagger}, \tag{A5}$$

where $K_{\mathbf{n}}^{\nu}$ is the νth-order commutator of the operator $V_{\mathbf{n}}$ and $H_{\mathbf{n}}$, and

$$G_{\mathbf{nm}}^{\nu}(t)=-i\theta(t)\langle\exp(iH_{R}t)K_{\mathbf{n}}^{\nu}B_{\mathbf{n}}\exp(-iHt)B_{\mathbf{n}}^{\dagger}V_{\mathbf{m}}^{\dagger}\rangle, \tag{A6}$$

$$G_{\mathbf{n,lm}}^{\nu}(t)=-i\theta(t)\langle\exp(iH_{R}t)\widetilde{K}_{\mathbf{n}}^{\nu}B_{\mathbf{l}}V_{\mathbf{l}}\times\exp(-iHt)B_{\mathbf{m}}^{\dagger}V_{\mathbf{m}}^{\dagger}. \tag{A7}$$

Equation (A4) is nonlinear because of the last term on the right-hand side, where in comparison to the function being differentiated the power of the operator V has increased by unity (cf. (A6) and (A7)).

Now, if in the terms of the chain we go over to the Fourier integral and substitute each $G_{\mathbf{nm}}^{\nu}(E)$ into the previous equation, we arrive at an expression for the Green's function in the approximation of the first iteration cycle:

$$G_{\mathbf{nm}}^{\nu}(E+i\gamma)=\sum_{\nu=0}^{\infty}\frac{1}{(E-\varepsilon^{0}+i\gamma)^{\nu+1}}\left[\langle\widetilde{K}_{\mathbf{n}}^{\nu}\rangle\delta_{\mathbf{nm}}+\sum_{\mathbf{l}}{}'M_{\mathbf{nl}}^{0}G_{\mathbf{n,lm}}^{\nu}(E+i\gamma)\right]. \tag{A8}$$

The next step consists in obtaining a similar chain for the function $G_{\mathbf{n,lm}}^{\nu}$. The result of the second iteration cycle is

$$G_{\mathbf{n,lm}}^{\nu}(E+i\gamma)=\sum_{\nu_1=0}^{\infty}\frac{1}{(E-\varepsilon^{0}+i\gamma)^{\nu_1+1}}\left[\langle\widetilde{K}_{\mathbf{nm}}^{\nu\nu_1}\rangle\delta_{\mathbf{lm}}+\sum_{\mathbf{p}}{}'M_{\mathbf{lp}}^{0}G_{\mathbf{nl,pm}}^{\nu\nu_1}(E+i\gamma)\right]. \tag{A9}$$

Here

$$\widetilde{K}_{\mathbf{nm}}^{\nu\nu_1}=K_{\mathbf{nm}}^{\nu\nu_1}V_{\mathbf{m}}^{\dagger}, \tag{A10}$$

where $K_{\mathbf{nm}}^{\nu\nu_1}$ is the νth-order commutator of the product $\widetilde{K}_{\mathbf{n}}^{\nu}V_{\mathbf{m}}$ and the operator $H_{\mathbf{m}}$ specified in (A2), and $G_{\mathbf{nl,pm}}^{\nu\nu_1}(E+i\gamma)$ is the Fourier transform of the function

$$G_{\mathbf{nl,pm}}^{\nu\nu_1}(t)=-i\theta(t)\langle\exp(iH_{R}t)\widetilde{K}_{\mathbf{nl}}^{\nu\nu_1}B_{\mathbf{p}}V_{\mathbf{p}}\times\exp(-iHt)B_{\mathbf{m}}^{\dagger}V_{\mathbf{m}}^{\dagger}\rangle. \tag{A11}$$

The equation for (A11) and all the subsequent cycles are constructed by analogy with (A8) and (A9).

Below, in examining commutators of the type (A5) and (A10), it will be convenient to employ the following representation:

$$K_{\mathbf{n}}^{\nu}=\sum_{p=0}^{\nu}C_{\nu}^{p}(-1)^{p}(H_{\mathbf{n}})^{p}V_{\mathbf{n}}(H_{\mathbf{n}})^{\nu-p}=\frac{\nu!}{2\pi i}\oint_{|z|=0}\frac{\exp(-zH_{\mathbf{n}})V_{\mathbf{n}}\exp(zH_{\mathbf{n}})}{z^{\nu+1}}\,dz, \tag{A12}$$

where C_{ν}^{p} are binomial coefficients.

Using the representation (A12) and the commutativity of $a_{\mathbf{ns}}^{\dagger},a_{\mathbf{ns}}$ and $b_{\mathbf{k}r}^{\dagger},b_{\mathbf{k}r}$, we can write a commutator of arbitrary order as follows:

$$\widetilde{K}_{\mathbf{n}_1,\ldots,\mathbf{n}_j}^{\nu_1,\ldots,\nu_j}=\prod_{l=1}^{j}\frac{\nu_l!}{2\pi i}\oint_{|z_l|=0}\frac{dz_l}{z_l^{\nu+1}}K_{v}(z_1,\ldots,z_j)\times K_{R}(z_l,\ldots,z_j), \tag{A13}$$

$$K_v(z_1,\ldots,z_j)=\prod_{p=1}^{j}\exp(-z_pH^{\nu}_{\mathbf{n}_p})\times T_q\prod_{q=1}^{j}P_{\mathbf{n}_q}\exp(z_qH^{\nu}_{\mathbf{n}_q})P^{\dagger}_{\mathbf{n}_q}, \tag{A14}$$

$$K_R(z_1,\ldots,z_j)=\prod_{p=1}^{j}\exp(-z_pH_R)\times T_q\prod_{q=1}^{j}F_{\mathbf{n}_q}\exp(z_qH_R)F^{\dagger}_{\mathbf{n}_q}. \tag{A15}$$

Here T_q stands for the product ordered in index q of the group of operators under the product sign. Averaging (A15) via the density matrix

$$\rho=\frac{\exp(-H_R/kT)}{\mathrm{Tr}[\exp(-H_R/kT)]} \tag{A16}$$

and noting that

$$\exp(-zH_R)\exp(D^{\mathbf{k}r}_{\mathbf{n}}b_{\mathbf{k}r})\exp\, z(H_R)=\exp[D^{\mathbf{k}r}_{\mathbf{n}}\exp(\omega_{\mathbf{k}r}z)b_{\mathbf{k}r}], \tag{A17}$$

$$\times\exp(-zH_R)\exp[-(D^{\mathbf{k}r}_{\mathbf{n}})^{*}b^{\dagger}_{\mathbf{k}r}]\exp(zH_R)=\exp[-(D^{\mathbf{k}r}_{\mathbf{n}})^{*}\exp(-\omega_{\mathbf{k}r}z)b^{\dagger}_{\mathbf{k}r}], \tag{A18}$$

we obtain

$$\langle K_R(z_1,\ldots,z_j)\rangle=\prod_{\mathbf{k}r}\prod_{q=1}^{j}\exp\left\{\left|D_{\mathbf{k}r}\right|^2\left[(\exp(\omega_{\mathbf{k}r}z_q)-1)+\left(\exp\frac{\omega_{\mathbf{k}r}z_q}{2}-\exp\left(-\frac{\omega_{\mathbf{k}r}z_q}{2}\right)\right)^2\langle n_{\mathbf{k}r}\rangle\right]\right\}\times\exp\left\{\left|D_{\mathbf{k}r}\right|^2\sum_{l=1}^{j-1}\sum_{i>l}^{j}[f^{\mathbf{k}r}_{\mathbf{n}_l\mathbf{n}_i}(1+\langle n_{\mathbf{k}r}\rangle)+\tilde{f}^{\,\mathbf{k}r}_{\mathbf{n}_l\mathbf{n}_i}\langle n_{\mathbf{k}r}\rangle]\right\}, \tag{A19}$$

where

$$f^{\,\mathbf{k}r}_{\mathbf{n}_l\mathbf{n}_i}=\exp[i\mathbf{k}\cdot(\mathbf{n}_l-\mathbf{n}_i)][\exp(\omega_{\mathbf{k}r}z_l)-1]\times[\exp(\omega_{\mathbf{k}r}z_i)-1]\prod_{l<q<i}\exp(\omega_{\mathbf{k}r}z_q), \tag{A20}$$

$$\tilde{f}^{\,\mathbf{k}r}_{\mathbf{n}_l\mathbf{n}_i}=\exp[-i\mathbf{k}\cdot(\mathbf{n}_l-\mathbf{n}_i)][1-\exp(-\omega_{\mathbf{k}r}z_l)]\times[1-\exp(-\omega_{\mathbf{k}r}z_i)]\prod_{l<q<i}\exp(-\omega_{\mathbf{k}r}z_q), \tag{A21}$$

and $\langle n_{\mathbf{k}r}\rangle=[\exp(\omega_{\mathbf{k}r}/kT)-1]^{-1}$ is the average number of lattice phonons with quantum numbers $n_{\mathbf{k}r}$ at the temperature T.

If in the vibronic subsystem we ignore the quadratic interaction, then for (A19) we can easily obtain an expression for the expectation value of the commutator (A14). Replacing $|D_{\mathbf{k}r}|^2$ by α_s^2 and averaging over the ground state of the subsystem, we obtain

$$\langle K_v(z_1,\ldots,z_j)\rangle=\prod_{s}\prod_{q=1}^{j}\exp\{\alpha_s^2[\exp(\omega_s^0z_q)-1]\}\times\prod_{l\neq p}(1+\delta_{\mathbf{n}_l\mathbf{n}_p}f^{s}_{lp}), \tag{A22}$$

$$f^{s}_{lp}=\exp\left\{\alpha_s^2[\exp(\omega_s^0z_l)-1]\times[\exp(\omega_s^0z_p)-1]\prod_{i\neq l,p}\exp(\omega_s^0z_i)\right\}-1. \tag{A23}$$

Here the index i assumes only those values that lie in the interval between the largest and the smallest among the pair of numbers l and p. Bearing in mind that

$$\exp(-zH^{v}_{\mathbf{n}})(a_{\mathbf{n}s})^p\exp(zH^{v}_{\mathbf{n}})=[a_{\mathbf{n}s}\exp(\omega^f_sz)]^p, \tag{A24}$$

$$\exp(-zH^{v}_{\mathbf{n}})(a^{\dagger}_{\mathbf{n}s})^p\exp(zH^{v}_{\mathbf{n}})=[a^{\dagger}_{\mathbf{n}s}\exp(-\omega^f_sz)]^p, \tag{A25}$$

and

$$\exp(\alpha_sa_{\mathbf{n}s})\exp(\beta_sa^{\dagger}_{\mathbf{m}s})=\exp(\beta_sa^{\dagger}_{\mathbf{m}s})\exp(\alpha_sa_{\mathbf{n}s})\times\{1+\delta_{\mathbf{nm}}[\exp(\alpha_s\beta_s)-1]\}, \tag{A26}$$

we can obtain formula (A22) directly from (A14), provided that first we order the exponential operators via (A26).

The structure of (A22) does not change if we allow for the frequency defect $\Delta\omega$, but formula (A23) becomes extremely cumbersome.

APPENDIX B: SPLITTING THE CORRELATION FUNCTIONS

To split the correlation functions, it is convenient to assume that the order of a correlator is the number of groups of cofactors numbered by the same index of an ordered product. Thus, the first term of the series (10) contains a first-order correlator, the second term a second-order correlator, etc. If we now replace the expectation values of (A14) and (A15) by the products of the expectation values of the exponential operators

$$\exp(-z_qH^{v}_{\mathbf{n}_q})P_{\mathbf{n}_q}\exp(z_qH^{v}_{\mathbf{n}_q})P^{\dagger}_{\mathbf{n}_q}$$

and

$$\exp(-z_qH_R)F_{\mathbf{n}_q}\exp(z_qH_R)F^{\dagger}_{\mathbf{n}_q},$$

respectively, the correlator of the jth order can be factorized, with the result that

$$\langle\tilde{K}^{\nu_1\cdots\nu_j}_{\mathbf{n}_1\cdots\mathbf{n}_j}\rangle\approx\prod_{q=1}^{j}\langle\tilde{K}^{\nu_q}_{\mathbf{n}_q}\rangle. \tag{A27}$$

This approximation is equivalent to allowing in (A19) and (A22) for coordinate-independent terms.

To estimate the contribution of the terms that depend on the coordinates, we examine the terms in the correlator (A22) that are proportional to $\delta_{\mathbf{n}_l\mathbf{n}_p}$. To simplify matters, we consider only one branch of vibrations.

Noting that coupling in the resonant interaction is forbidden for operators belonging to the same lattice site $\mathbf{n}_l=\mathbf{n}_p$, for each correlator we can easily calculate the number of possible convolutions with the same structure. Convolutions obtained by coupling operators with equal (in absolute value) differences $|p-l|$ will be identical in structure. The third-order correlator has one convolution with $|p-l|_{\min}=2$, the fourth-order correlator has two such convolutions, the fifth-order correlator has three such convolutions, and so on. The fourth-order correlator has one convolution with $|p-l|=3$, the fifth-order correlator has two such convolutions, so that in each subsequent correlator the number of convolutions of this type is also increased by one. Clearly, the number of admissible pairings with any one of the differences $\Delta_{pl}=|p-l|$ in the jth-order correlator is $j-\Delta_{pl}$.

Now we can write in explicit form the contribution of the convolutions of the specified type to the sum (10). For a crystal with simple cubic symmetry with one molecule per unit cell we obtain (with allowance for the Franck–Condon transition)

$$G'_{\mathbf{nm}}(E)=\sum_{j=3}^{\infty} g_j, \tag{A28}$$

where

$$g_3=\delta_{\mathbf{nm}}K_3=\frac{1}{N}\sum_{\mathbf{Q}} \exp\{i\mathbf{Q}\cdot(\mathbf{n}-\mathbf{m})\}K_3,$$

$$g_4=\frac{1}{N}\sum_{\mathbf{Q}} \exp\{i\mathbf{Q}\cdot(\mathbf{n}-\mathbf{m})\}(K_4+2\varepsilon_{\mathbf{Q}}S_EK_3),\ldots,$$

$$g_\nu=\frac{1}{N}\sum_{\mathbf{Q}} \exp\{i\mathbf{Q}\cdot(\mathbf{n}-\mathbf{m})\} \times\sum_{p=1}^{\nu-2} p(\varepsilon_{\mathbf{Q}}S_E)^{p-1}K_{\nu-p+1},\ldots, \tag{A29}$$

$$K_r=\frac{1}{N}\sum_{\mathbf{Q}'} (\varepsilon_{\mathbf{Q}'})^{r-1}\sum_{l=1}^{\infty} \frac{\alpha^{2l}}{l!}(B_l)^2(S_l)^{r-1}. \tag{A30}$$

Here

$$B_l=\sum_{p=0}^{l} C_l^p(-1)^{l-p}S_p, \tag{A31}$$

$$S_p=\exp(-\alpha^2)\sum_{q=0}^{\infty} \frac{\alpha^{2q}}{q!}\frac{1}{E-\varepsilon^0-(p+q)\omega^0+i\gamma}, \tag{A32}$$

$$S_E=S_{p=0}. \tag{A33}$$

In (A29) we first collect terms containing K_r with equal values of r and then sum the resulting subsequences over index r. As a result we get

$$G'_{\mathbf{nm}}(E)=\frac{1}{N}\sum_{\mathbf{Q}} \exp\{i\mathbf{Q}\cdot(\mathbf{n}-\mathbf{m})\}G'_{\mathbf{Q}}(E), \tag{A34}$$

$$G'_{\mathbf{Q}}(E)=\frac{1}{(1-\varepsilon_{\mathbf{Q}}S_E)^2}\sum_{l=1}^{\infty} \frac{\alpha^{2l}}{l!}(B_l)^2\frac{1}{N}\sum_{\mathbf{Q}'} \frac{\varepsilon_{\mathbf{Q}'}^2S_l}{1-\varepsilon_{\mathbf{Q}'}S_l}. \tag{A35}$$

The expression (35) can be shown to formally split into groups of terms of the type

$$\begin{aligned} G'_{\mathbf{Q}}(E)=\sum \cdots\Bigg[&\frac{1}{\varepsilon_s-\varepsilon_s^l}\frac{1}{(E-\varepsilon^0-\varepsilon_s+i\gamma)^2} \\ &-\frac{1}{(\varepsilon_s-\varepsilon_s^l)^2}\Bigg(\frac{1}{E-\varepsilon^0-\varepsilon_s+i\gamma} \\ &-\frac{1}{E-\varepsilon^0-\varepsilon_s^l+i\gamma}\Bigg)\Bigg]+\cdots \\ &+\Bigg[\frac{1}{\varepsilon_s-(p+q)\omega^0}\frac{1}{[E-\varepsilon^0-(p+q)\omega^0+i\gamma]^2} \\ &-\frac{1}{[\varepsilon_s-(p+q)\omega^0]^2}\Bigg(\frac{1}{E-\varepsilon^0-\varepsilon_s+i\gamma} \\ &-\frac{1}{E-\varepsilon^0-(p+q)\omega^0+i\gamma}\Bigg)\Bigg], \end{aligned} \tag{A36}$$

with ε_s and ε_s^l the roots of the equations $1-\varepsilon_{\mathbf{Q}}\,\mathrm{Re}\,S_E=0$ and $1-\varepsilon_{\mathbf{Q}'}\,\mathrm{Re}\,S_l=0$, respectively.

Obviously,

$$\int_{-\infty}^{\infty} \mathrm{Im}\ G'_{\mathbf{Q}}(E)dE=0. \tag{A37}$$

Since $G'_{\mathbf{Q}}(E)$ is an analytic function of a complex variable, formula (A36) allows for a simple hydrodynamic analogy:[37] it describes potential flow of an incompressible liquid whose complex-valued velocity is initiated by a system of dipoles (the first terms in the square brackets) and by doublets of point sources (the second terms) with strengths equal in absolute value but opposite in sign. In this interpretation the function $G_{\mathbf{Q}\mu}(E)$ specified by (14) describes liquid flow initiated by a system of isolated point sources.

[1] V. M. Agranovich, *Theory of Excitons* [in Russian], Nauka, Moscow (1968).
[2] A. S. Davydov, *Theory of Molecular Excitations*, Plenum Press, New York (1971).
[3] V. M. Agranovich and M. D. Galanin, *Electronic Excitation Energy Transfer in Condensed Matter*, North-Holland, Amsterdam (1982).
[4] V. L. Broude, E. I. Rashba, and E. F. Sheka, *Spectroscopy of Molecular Excitons*, Springer, Berlin (1985).
[5] K. K. Rebane, *Elementary Theory of the Vibrational Structure of the Spectra of Impurity Crystals* [in Russian], Nauka, Moscow (1979).
[6] I. A. Malkin and V. I. Man'ko, *Dynamic Symmetries and Coherent States of Quantum Systems* [in Russian], Nauka, Moscow (1979).
[7] I. S. Osad'ko, Usp. Fiz. Nauk **128**, 31 (1979) [Sov. Phys. Usp. **22**, 311 (1979)].
[8] E. F. Sheka and I. P. Terenetskaya, Chem. Phys. **8**, 996 (1975).
[9] S. L. Robinette, S. H. Stevenson, and G. J. Smoll, J. Chem. Phys. **69**, 5231 (1978).
[10] V. L. Broude, Preprint of the Solid-State Physics Institute of the USSR Academy of Sciences, Chernogolovka (1973).
[11] J. Tokura and T. Koda, J. Lumin. **18/19**, 467 (1979).
[12] D. P. Craig and S. H. Walmsley, Mol. Phys. **4**, 13 (1968).
[13] D. W. Schlosser and M. R. Philpott, J. Chem. Phys. **77**, 1969 (1982).
[14] D. P. Craig and S. H. Walmsley, in *Physics and Chemistry of the Organic*

Solid State, Vol. 1, D. Fox, M. M. Labes and A. Weissberger (Eds.), Wiley, New York (1965).
[15] I. V. Brovchenko, V. I. Tovstenko, and M. T. Shpak, Fiz. Tverd. Tela (Leningrad) **31**, No. 12, 1 (1989) [Sov. Phys. Solid State **31**, 2027 (1989)].
[16] S. A. Rice, J. Chem. Phys. **42**, 2948 (1965).
[17] M. S. Soskin, Ukr. Fiz. Zh. (Russ. Ed.) **5**, 507 (1960).
[18] I. P. Terenetskaya, Zh. Org. Soed. **36**, 516 (1974).
[19] N. I. Ostapenko, M. P. Chernomorets, and M. T. Shpak, Ukr. Fiz. Zh. (Russ. Ed.) **21**, 272 (1976).
[20] É. I. Rashba, Zh. Éksp. Teor. Fiz. **54**, 542 (1968) [Sov. Phys. JETP **27**, 292 (1968)].
[21] A. A. Serikov, Phys. Status Solidi B **44**, 733 (1971).
[22] M. Cini and M. Cuozzo, J. Phys.: Condens. Matter **8**, 9011 (1996).
[23] Y. Zhao, D. W. Brown, and K. Lindenberg, J. Chem. Phys. **100**, 2334 (1994).
[24] C. Qinghu, F. Minghu, Z. Qirui *et al.*, J. Phys.: Condens. Matter **8**, 7139 (1996).
[25] C. Qinghu, W. Kelin, and W. Shaolong, Phys. Rev. B **50**, 14583 (1994).
[26] P. Petelenz, Chem. Phys. Lett. **229**, 439 (1994).
[27] P. Petelenz, Chem. Phys. Lett. **240**, 605 (1995).
[28] J. Appel, ''Polarons,'' in *Solid State Physics: Advances in Research and Applications*, H. Ehrenreich, F. Seitz, and D. Turnbull (Eds.), Vol. 21, Academic Press, New York (1968); Yu. A. Firsov, in *Polarons* [in Russian], Yu. A. Firsov (ed.), Nauka, Moscow (1975).
[29] F. Duschinsky, Acta Physicochim. URSS **2**, 551 (1937).
[30] N. I. Ostapenko, V. I. Sugakov, M. P. Chernomorets *et al.*, Phys. Status Solidi B **93**, 493 (1979).
[31] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[32] I. M. Lifshitz, Zh. Éksp. Teor. Fiz. **17**, 1017 (1947); **17**, 1076 (1947).
[33] A. M. Kosevich, *The Basics of Crystal Lattice Mechanics* [in Russian], Nauka, Moscow (1972).
[34] Yu. A. Izyumov and M. V. Medvedev, *Magnetically Ordered Crystals Containing Impurities*, Consultants Bureau, New York (1973).
[35] É. L. Bokhenkov, I. Natkanets, and E. F. Sheka, Zh. Éksp. Teor. Fiz. **70**, 1027 (1976) [Sov. Phys. JETP **43**, 536 (1976)].
[36] M. Suzuki, T. Yokyama, and M. Ito, Spectrochim. Acta A **34**, 109 (1968).
[37] M. A. Lavrent'ev and B. V. Shabat, *Problems of Hydrodynamics and Their Mathematical Models* [in Russian], Nauka, Moscow (1977).

Translated by Eugene Yankovsky

Inelastic light scattering by electron excitations with large wave vectors in a 2D magnetoplasma

A. O. Govorov

Institute of Semiconductor Physics, Russian Academy of Sciences Siberian Branch, 630090, Novosibirsk-90, Russia; Sektion Physik der Ludwig-Maximilians-Universität, 80539, München, Germany
(Submitted 27 November 1996)
Zh. Éksp. Teor. Fiz. **112**, 1041–1054 (September 1997)

Microscopic mechanisms of inelastic light scattering in an interacting electron plasma in semiconductor heterostructures are considered. In the dipole limit, the cross section consists of two main contributions: the first is related to a disorder-induced mechanism and the second arises from the Coulomb interaction. The spectra of disorder-induced light scattering are described in terms of correlation functions of a random potential. The spectrum induced by the Coulomb interaction arises from two-quasiparticle excitations. The mechanisms which are studied in this paper result in the appearance of large wave vector excitations in the spectra of resonant light scattering. These results can be used to model the experimentally observed appearance of the roton density of states in light scattering spectra in the integer quantum Hall regime of a two-dimensional system. Furthermore, we show that the lineshape of spectra strongly depends on the character of disorder and, in particular, on the spatial positions of impurities with respect to a quantum well. © *1997 American Institute of Physics.* [S1063-7761(97)01809-X]

1. INTRODUCTION

Raman studies of two-dimensional (2D) systems in a high magnetic field are currently an active area of research.[1] In particular, inelastic light scattering (LS) permits the observation of the roton excitations in the regimes of the integer and fractional quantum Hall effects (QHE's).[2–4] The magnetoroton excitations in the integer QHE regime have characteristic wave vectors of the order of $1/l_c$, where l_c is the magnetic length. At the same time, in-plane momentum transfers of the order of $1/l_c$ are not easily accessible in experiments. Magnetorotons in LS spectra were interpreted as arising from breakdown of momentum conservation in the presence of residual disorder.[1,2]

The feature of the experiments mentioned above is that LS spectra contain excitations with relatively large wave vectors. In this paper, we consider specific mechanisms of LS, which allow the observation of such excitations.

The effect of disorder on LS spectra in the integer QHE regime was investigated theoretically in Ref. 5 in the framework of a phenomenological approach. The cross section, calculated in Ref. 5, reflects the density of states of electron excitations and is written as

$$\frac{d^2\sigma}{d\Omega d\omega} \propto \int S_0(\omega,q) f(q) d^2q, \tag{1}$$

where ω is the photon energy transfer in the LS process, q is the in-plane wave vector of an excitation, and $S_0(\omega,q)$ is the structure factor of the system. The function $f(q)$ describes breakdown of momentum conservation and is written in Ref. 5 in the Lorentzian form: $f(q)=(\alpha/\pi)/(q^2+\alpha^2)$, where α is the phenomenological broadening parameter.

In the fractional QHE regime, LS was studied theoretically in a work of Platzman and Song He,[6] where the authors have obtained numerical data for the intra-Landau-level spectra. They considered LS as a shake-up process, which results in two-excitation lines. This process is directly connected with the Coulomb interaction and was described in Ref. 6 by phenomenological matrix elements. Shake-up processes for the case of two-phonon LS from the Wigner crystal were considered in Ref. 7.

In the present paper, we study resonant LS in the dipole limit ($k\to 0$, where k is the light momentum transfer). To calculate the LS spectra, we find the Hamiltonian responsible for dipole-allowed LS by using a general formalism developed in Ref. 8. The cross section is expanded into a series of the parameter $1/(E_g-\omega_1)$, where E_g is the optical-gap energy, and ω_1 is the laser frequency. We obtained analytical expressions for the amplitudes of LS induced by disorder and by the Coulomb interaction. The cross section of disorder-induced LS is expressed in terms of correlation functions of a random potential, which determines the characteristic wave vectors of excitations in Raman spectra. The Coulomb interaction in an electron system results in two-quasiparticle excitations in LS spectra. The mechanisms of LS, which we consider here, were studied earlier in Refs. 5 and 6 by using various phenomenological approaches. In this paper, we develop from first principles a theory describing such LS mechanisms. The results obtained by us can be applied both to bulk and 2D systems. We focus on 2D semiconductor heterostructures because LS mechanisms, which involve magnetoexcitations with large wave vectors, play the most important role in these systems. In particular, we calculate the LS spectra of a 2D magnetoplasma with the filling factor $\nu=2$.

2. SYSTEM RESONANT LIGHT SCATTERING IN AN ELECTRON

Resonant LS is connected with two virtual interband processes of absorption and emission which are induced by

the incident and scattered photons, respectively. In the following, we will consider resonance between the lowest 2D subbands in the conduction and heavy-hole valence bands. Resonances with the light-hole and split-off valence bands can be described similarly. In addition, we assume that electrons occupy only the lowest 2D subband in the conduction band.

The cross section of LS and the structure factor $S(\omega)$ are given by (see Ref. 8)

$$\frac{d^2\sigma}{d\Omega d\omega}=\frac{\omega_2}{\omega_1}\frac{e^4}{c^4 m_0^4}S(\omega),$$

$$S(\omega)=\sum_F |\langle F|\hat{V}_{\rm eff}|0\rangle|^2\delta(E_0-E_F+\omega)$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\langle 0|\hat{V}_{\rm eff}^{+}\hat{V}_{\rm eff}(t)|0\rangle e^{i\omega t}dt, \qquad (2)$$

where $|0\rangle$ and $|F\rangle$ are the initial and final states of the many-electron system, E_0 and E_F are their energies, and m_0 is the free-electron mass; the operator $\hat{V}_{\rm eff}$ is the interaction Hamiltonian describing LS and

$$\hat{V}_{\rm eff}(t)=\exp(-i\hat{H}_{\rm tot}t)\hat{V}_{\rm eff}\exp(i\hat{H}_{\rm tot}t),$$

where $\hat{H}_{\rm tot}$ is the Hamiltonian of the crystal; $\omega_{1(2)}$ are energies of incident (scattered) photons, $\omega=\omega_1-\omega_2$ is the energy transfer, and $\hbar=1$. The temperature is assumed to be zero. Assuming resonance, the matrix elements $\langle F|\hat{V}_{\rm eff}|0\rangle$ are (see Ref. 8)

$$\langle F|\hat{V}_{\rm eff}|0\rangle=\sum_N\frac{\langle F|\hat{j}_2|N\rangle\langle N|\hat{j}_1|0\rangle}{\omega_1+E_0-E_N}$$

$$=-i\left\langle F\left|\int_0^{\infty}\hat{j}_2\hat{j}_1(t)e^{i\omega_1 t}dt\right|0\right\rangle, \qquad (3)$$

where $|N\rangle$ are intermediate many-electron states. The operators $\hat{j}_1$ and $\hat{j}_2$ describe the interband optical processes assisted by the incident and scattered photons, respectively. The states $|N\rangle$ are characterized by a single hole in the valence band and one additional electron in the conduction band.

Single-electron states of the conduction band are given by $|\alpha\rangle=e^{i\mathbf{pr}}\phi_c(z)|\sigma\rangle$, where $\mathbf{r}$ and z are the in-plane and normal coordinates, respectively; $\mathbf{p}$ is the electron momentum, $\phi_c(z)$ is the wave function describing size-quantization and $\sigma=\pm 1/2$ are the spin indexes. Single-electron states of the valence band can be written similarly: $|\gamma\rangle=e^{i\mathbf{pr}}\phi_v(z)|\mathbf{J}\rangle$, where $\phi_v(z)$ is the wave function connected with size-quantization, and $\mathbf{J}$ is the angular momentum of heavy holes. For simplicity, we disregard mixing between the valence bands.

The Hamiltonian of the electron system is $\hat{H}_{\rm tot}=\hat{H}_b+E_c\hat{n}_c+E_v\hat{n}_v$, where the subscripts c and v refer to the first subbands in the conduction and valence bands, respectively; $\hat{H}_b$ is the Hamiltonian describing intraband energies and direct Coulomb interaction between electrons of different bands, $\hat{n}_c(\hat{n}_v)$ are the operators of electron numbers in the conduction (valence) bands, and $E_c(E_v)$ are the energies of electrons at zero in-plane momentum in corresponding 2D subbands. The operator $\hat{H}_b=\hat{H}_c+\hat{H}_v+\hat{H}_{cv}^{\rm int}$, where $\hat{H}_{cv}^{\rm int}$ is the operator of direct Coulomb interaction between electrons of the conduction and valence bands. Exchange interaction between electrons of different bands is ignored. The operators $\hat{H}_c$ and $\hat{H}_v$ describe the intraband energies of electrons: $\hat{H}_c=\hat{T}_c+\hat{u}_c^{\rm def}+\hat{H}_{cc}^{\rm int}$ and $\hat{H}_v=\hat{T}_v+\hat{u}_v^{\rm def}$. Here $\hat{T}_c$ and $\hat{T}_v$ are the operators of kinetic energy; $u_c^{\rm def}(\mathbf{r})$ and $u_v^{\rm def}(\mathbf{r})$ are random potentials in the conduction and valence bands, respectively; and $\hat{H}_{cc}^{\rm int}$ is the operator of the Coulomb interaction in the conduction band. Here we include in the Hamiltonian $\hat{H}_{\rm tot}$ only the terms related to the conduction and valence subbands, which are coupled by interband resonant transitions. The Coulomb potentials are given by the matrix elements:

$$U_{cc}(\mathbf{r}-\mathbf{r}')=\langle\phi_c(z)\phi_c(z')|U\times(\sqrt{(z-z')^2+(\mathbf{r}-\mathbf{r}')^2})|\phi_c(z)\phi_c(z')\rangle,$$

$$U_{cv}(\mathbf{r}-\mathbf{r}')=\langle\phi_c(z)\phi_v(z')|U\times(\sqrt{(z-z')^2+(\mathbf{r}-\mathbf{r}')^2})|\phi_c(z)\phi_v(z')\rangle, \qquad (4)$$

where $U(R)=e^2/\epsilon R$, and ϵ is the dielectric constant.

A method to simplify the operator $\hat{V}_{\rm eff}$ was proposed in Refs. 7 and 9. According to this method, we find the commutator $[E_c\hat{n}_c+E_v\hat{n}_v,\hat{j}_1]=(E_c-E_v)\hat{j}_1=E_g\hat{j}_1$. At the same time, the operator $\hat{H}_b$ does not change the number of particles in any band; i.e., the commutator $[E_c\hat{n}_c+E_v\hat{n}_v,\hat{H}_b]=0$. Thus, the effective Hamiltonian of resonant LS can be written as

$$\hat{V}_{\rm eff}=-i\int_0^{\infty}\hat{j}_2\hat{j}_1(t)\exp(i\omega_1 t)dt=-i\int_0^{\infty}\hat{j}_2\times\exp(-i\hat{H}_b t)\hat{j}_1\exp(i\hat{H}_b t)\exp[i(\omega_1-E_g)t]dt. \qquad (5)$$

We now expand Eq. (5) in a series of the operator $\hat{H}_b$ using the equation

$$e^{\hat{a}}\hat{b}e^{-\hat{a}}=\hat{b}+\frac{[\hat{a},\hat{b}]}{1!}+\frac{[\hat{a},[\hat{a},\hat{b}]]}{2!}+\dots .$$

After integration in Eq. (5), we have

$$\hat{V}_{\rm eff}=\hat{V}_1+\hat{V}_2+\hat{V}_3+\dots,$$

$$\hat{V}_1=\frac{\hat{j}_2\hat{j}_1}{\Delta},\quad \hat{V}_2=\frac{\hat{j}_2[\hat{H}_b,\hat{j}_1]}{\Delta^2}, \qquad (6)$$

$$\hat{V}_3=\frac{\hat{j}_2[\hat{H}_b,[\hat{H}_b,\hat{j}_1]]}{\Delta^3},\dots,$$

where $\Delta=\omega_1-E_g$. The expansion (6) is valid if the value Δ is much larger than the energies of electron excitations in a plasma; i.e., $|\Delta|\gg\epsilon_k$, ϵ_C, Γ, where ϵ_k and ϵ_C are the characteristic kinetic and Coulomb energies, and Γ is the broadening of electron levels due to a random potential. In a high magnetic field, $\epsilon_k\sim\omega_c$ and $\epsilon_C\sim e^2/(\epsilon l_c)$, where ω_c is the

cyclotron energy. At the same time, we assume $|\Delta| \ll E_g$. Below we will focus only on two first terms in the expansion (6).

2.1. The operator $\hat{V}_1$

The first term in Eq. (6) was calculated by Hamilton and McWhorter.[8] In the Kane model, the operator $\hat{V}_1$ is written as (see Ref. 8)

$$\hat{V}_1 = -\frac{1}{\Delta}\,[f_e \hat{\rho}_e(\mathbf{k}_\parallel) + f_s \rho_s(\mathbf{k}_\parallel)], \tag{7}$$

$$\hat{\rho}_e(\mathbf{k}_\parallel) = \sum_i \exp(i\mathbf{k}_\parallel \mathbf{r}_i), \quad \hat{\rho}_s(\mathbf{k}_\parallel) = \sum_i \exp(i\mathbf{k}_\parallel \mathbf{r}_i)\hat{\sigma}_{iz}, \tag{8}$$

where $\mathbf{k}_\parallel = \mathbf{k}_{1\parallel} - \mathbf{k}_{2\parallel}$, $\mathbf{k}_1$ and $\mathbf{k}_2$ are the wave vectors of incident and scattered photons, respectively; $\mathbf{r}_i$ is the in-plane coordinate of the i-electron, $\hat{\sigma}_{iz}$ is the Pauli matrix, and $\hat{\rho}_e$ and $\hat{\rho}_s$ are the operators of electron and spin densities, respectively. The parameters $f_{e(s)}$ in Eq. (7) show selection rules of LS in resonance with the heavy-hole valence band.[8,10] $f_e = D(\mathbf{e}_{1\parallel}\mathbf{e}^*_{2\parallel})$ and $f_s = iD[\mathbf{e}_1 \mathbf{e}^*_2]_z$, where $\mathbf{e}_1(\mathbf{e}_2)$ are the polarization vectors of incident (scattered) photons,

$$D = \frac{|P_{cv}|^2}{2}\,\langle \phi_v(z)|\exp(-ik_{2\perp}z)|\phi_c(z)\rangle \times \langle \phi_c(z)|\exp(ik_{1\perp}z)|\phi_v(z)\rangle,$$

and P_{cv} is the interband matrix element.

2.2. The operator $\hat{V}_2$

The second term in Eq. (6) can be written as

$$\hat{V}_2 = -\frac{1}{\Delta^2}\,[f_e \hat{Q}_e(\mathbf{k}) + f_s \hat{Q}_s(\mathbf{k}) + \hat{C}(\mathbf{q})],$$

$$\hat{Q}_e(\mathbf{k}) = \sum_i \left\{ \frac{\hat{\mathbf{p}}_i^2}{2m_c} + \frac{(\hat{\mathbf{p}}_i - \mathbf{k}_{1\parallel})^2}{2m_v} + u_c^{\mathrm{def}}(\mathbf{r}_i) - u_v^{\mathrm{def}}(\mathbf{r}_i) \right\} \exp(i\mathbf{k}_\parallel \mathbf{r}_i), \tag{9}$$

$$\hat{Q}_s(\mathbf{k}) = \sum_i \hat{\sigma}_{iz} \left\{ \frac{\hat{\mathbf{p}}_i^2}{2m_c} + \frac{(\hat{\mathbf{p}}_i - \mathbf{k}_{1\parallel})^2}{2m_v} + u_c^{\mathrm{def}}(\mathbf{r}_i) - u_v^{\mathrm{def}}(\mathbf{r}_i) \right\} \exp(i\mathbf{k}_\parallel \mathbf{r}_i),$$

where $m_{c(v)}$ are the effective masses in the conduction (valence) bands ($m_v > 0$). Here, the single-electron momentum $\hat{\mathbf{p}}_i$ should be written with allowance for the perpendicular magnetic field. The operator $\hat{C}(\mathbf{q})$ arises from the Coulomb interaction in the intermediate states. In the dipole limit, we have

$$\hat{C} = \sum_{\mathbf{q}} [U_{cc}(q) - U_{cv}(q)]\{f_e \hat{\rho}_e(\mathbf{q})\hat{\rho}_e(-\mathbf{q}) + f_s \hat{\rho}_s(\mathbf{q})\hat{\rho}_e(-\mathbf{q})\}. \tag{10}$$

The matrix elements $U_{cc}(q)$ and $U_{cv}(q)$ in Eq. (10) have opposite signs because the interband exciton in intermediate states is neutral. The value $U_{cc}(q) - U_{cv}(q)$ is nonzero if the wave functions $\phi_c(z)$ and $\phi_v(z)$ differ. Hence, the contribution $U_{cc} - U_{cv}$ can be essential in tilted quantum wells.

The operator $\hat{V}_1$ induces LS by charge- and spin-density excitations in corresponding geometries. The cross section connected with the operator $\hat{V}_1$ was calculated in a number of works for the case of bulk semiconductors[8,11,12] and for the case of quantum wells (for instance, see Ref. 13). It is essential that in Refs. 8 and 11–13 the cross section $d^2\sigma/d\Omega\, d\omega \to \delta(\omega)$ as $k \to 0$; i.e., the operator $\hat{V}_1$ leads to dipole-forbidden inelastic LS.

The operator $\hat{V}_2$ includes the contributions proportional to the operator of kinetic-energy density

$$\hat{T}_c(k_\parallel) = \sum_i \exp(i\mathbf{k}_\parallel \mathbf{r}_i)\,\frac{\hat{\mathbf{p}}_i^2}{2m_c}.$$

The operator $\hat{T}_c(k)$ results in so-called LS by fluctuations of kinetic-energy density, which was considered in Refs. 11, 12, and 14. In the approaches of Refs. 11, 12, and 14, the cross section induced by $\hat{T}_c(k)$ vanishes if $\omega \neq 0$ and $k = 0$. Therefore, the dipole-forbidden contributions, which can, in principle, play an important role in semiconductors, were taken into account in Refs. 11, 12, and 14.

2.3. Dipole-allowed inelastic light scattering

We now consider the operators $\hat{\rho}_e(\mathbf{k})$, $\hat{\rho}_s(\mathbf{k})$, $\hat{Q}_e(\mathbf{k})$, and $\hat{Q}_s(\mathbf{k})$ in the dipole limit $k \to 0$. At zero wave vector we have $\hat{\rho}_e(0) = \hat{n}_c$, $\hat{\rho}_s(0) = 2\hat{s}_z$ and, consequently, $[\hat{H}_c, \hat{n}_c] = 0$ and $[\hat{H}_c, \hat{s}_z] = 0$, where $\hat{s}_z$ is the total spin. These equalities mean that the operators $\hat{\rho}_e(0)$ and $\hat{\rho}_s(0)$ induce elastic LS.

Inelastic LS in the dipole limit can be connected with the operators $\hat{Q}_e(0)$ and $\hat{Q}_s(0)$, because the commutators $[\hat{Q}_e(0), \hat{H}_c]$ and $[\hat{Q}_s(0), \hat{H}_c]$ are nonzero. Usually, the Coulomb interaction and a random potential play the role of perturbations, i.e., $\epsilon_k \gg \epsilon_C, \Gamma$. Hence, it is convenient to exclude the kinetic energy from the operator $\hat{V}_2$. For instance, the operator $\hat{Q}_e(0)$ can be written as

$$\hat{Q}_e(0) = \sum_i \left\{ \frac{\hat{\mathbf{p}}_i^2}{2\mu} + u_c^{\mathrm{def}}(\mathbf{r}_i) - u_v^{\mathrm{def}}(\mathbf{r}_i) \right\} = \frac{m_c}{\mu}\,\hat{H}_c - \sum_i \left\{ \frac{m_c}{m_v}\, u_c^{\mathrm{def}}(\mathbf{r}_i) + u_v^{\mathrm{def}}(\mathbf{r}_i) \right\} - \frac{m_c}{\mu}\,\hat{H}_{cc}^{\mathrm{int}}, \tag{11}$$

where $1/\mu = 1/m_c + 1/m_v$. The operator $(m_c/\mu)\hat{H}_c$ in Eq. (11) contributes to the elastic LS process. Using similar transformations for all terms of $\hat{V}_2$, we have in the dipole limit $\hat{V}_1 + \hat{V}_2 = \hat{V}_{\mathrm{elas}} + \hat{W}$, where the operator $\hat{V}_{\mathrm{elas}}$ induces elastic LS. The contributions of kinetic energy remain now in the operator $\hat{V}_{\mathrm{elas}}$. Inelastic processes arise from the operator $\hat{W} = \hat{W}_{\mathrm{def}} + \hat{W}_{\mathrm{Coul}}$, where

$$\hat{W}_{\rm def}=\frac{1}{\Delta^2}\sum_i u_{\rm eff}(\mathbf{r}_i)[f_e+f_s\hat{\sigma}_{iz}]$$

$$=\frac{1}{\Delta^2}\sum_{\mathbf{q}} u_{\rm eff}(\mathbf{q})[f_e\hat{\rho}_e(\mathbf{q})+f_s\hat{\rho}_s(\mathbf{q})], \tag{12}$$

and

$$\hat{W}_{\rm Coul}=\frac{1}{\Delta^2}\sum_{i,i';i\neq i'}\left[\left(\frac{m_c}{2\mu}-1\right)U_{cc}(\mathbf{r}_i-\mathbf{r}_{i'})+U_{cv}(\mathbf{r}_i-\mathbf{r}_{i'})\right]\left(f_e+f_s\left(\frac{\hat{\sigma}_{iz}+\hat{\sigma}_{i'z}}{2}\right)\right)$$

$$=\frac{1}{\Delta^2}\sum_{\mathbf{q}}\left[\left(\frac{m_c}{2\mu}-1\right)U_{cc}(q)+U_{cv}(q)\right]\times\{f_e\hat{\rho}_e(\mathbf{q})\hat{\rho}_e(-\mathbf{q})+f_s\hat{\rho}_s(\mathbf{q})\hat{\rho}_e(-\mathbf{q})\}. \tag{13}$$

Here

$$u_{\rm eff}(\mathbf{r}_i)=u_v^{\rm def}(\mathbf{r}_i)+\frac{m_c}{m_v}u_c^{\rm def}(\mathbf{r}_i). \tag{14}$$

The functions $u_{\rm eff}(\mathbf{q})$ and $U_{cc}(\mathbf{q})$ are the Fourier transforms of the corresponding potentials. The operators $\hat{W}_{\rm def}$ and $\hat{W}_{\rm Coul}$ describe LS induced by disorder and by the Coulomb interaction, respectively. We note that the operator $\hat{W}$ includes terms proportional to the small parameters Γ and ϵ_C. The latter is essential because we may use first-order perturbation theory in this case (assuming $\epsilon_k\gg\Gamma,\ \epsilon_C$) to calculate the cross section.

3. LIGHT SCATTERING FROM THE 2D MAGNETOPLASMA

In this section, we intend to consider a 2D electron system in the perpendicular magnetic field B at the filling factor $\nu=2(\nu_\uparrow=\nu_\downarrow=1)$ and inter-Landau-level excitations with $\Delta l=1$ (Δl is the change of Landau-level number). We assume that the Coulomb energy is much less than the cyclotron frequency, $\epsilon_C=e^2/(\epsilon l_c)\ll\omega_c$, that allows us to use perturbation theory.[15,16] The creation operators for charge-density and spin-density excitations with the wave vector $\mathbf{q}$ are (see Ref. 15)

$$\hat{A}_e^+(\mathbf{q})=\frac{1}{\sqrt{N_s}}\hat{\rho}_e(\mathbf{q}),\quad \hat{A}_s^+(\mathbf{q})=\frac{1}{\sqrt{N_s}}\hat{\rho}_s(\mathbf{q}),$$

where N_e is the density of 2D electrons. The commutators with the Hamiltonian $\hat{H}_c$ are $[\hat{H}_c,\hat{A}_m^+(\mathbf{q})]=\omega_m\hat{A}_m^+(\mathbf{q})$, where the index $m=e(s)$ for charge-density (spin-density) excitations. The dispersions of magnetoexcitations (magnetoexcitons) ω_m, calculated in Refs. 15 and 16, are shown in the inset of Fig. 1. The wave functions of magnetoexcitons can now be written as $|\mathbf{q};m\rangle=\hat{A}_m^+(\mathbf{q})|0\rangle$.

The cross section can be written in terms of correlation functions

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\langle 0|\hat{A}_{m'}(\mathbf{q}')\exp(-i\hat{H}_ct)\hat{A}_m^+(\mathbf{q})\exp(i\hat{H}_ct)|0\rangle e^{i\omega t}dt$$

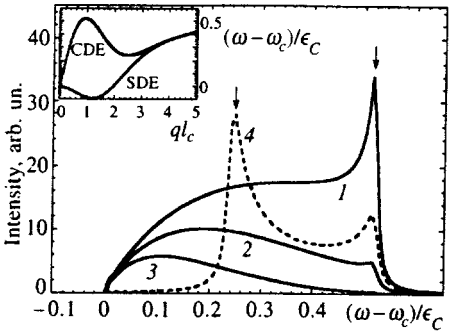

FIG. 1. Light scattering spectrum induced by disorder for the case of charge-density excitations, $\Gamma_e/\epsilon_C=0.01$; curves *1*, *2*, and *3* correspond to the impurity-induced mechanism with the parameters $z_0=200$, 300, and 500 Å, respectively; the magnetic length is $l_c=100$ Å. Curve *4* shows the spectrum of light scattering induced by a short-range random potential (interface defects). The arrows correspond to the critical points of the excitation dispersions. Inset: the dispersions of charge-density (CDE) and spin-density (SDE) excitations in the 2D electron plasma with the filling factor $\nu=2$ (see Ref. 16).

$$=-\frac{|L_{01}|^2}{\pi}\,\mathrm{Im}[G_m(\omega,p)]\,\delta_{m,m'}\delta_{\mathbf{q},\mathbf{q}'}, \tag{15}$$

where $\omega>0$, and

$$|L_{01}|^2=\frac{l_c^2q^2}{2}\exp\left(-\frac{l_c^2q^2}{2}\right).$$

The Green's function is

$$G_m(\omega,q)=\frac{1}{\omega-\omega_m(q)+i\Gamma_m(q)}, \tag{16}$$

where $1/\Gamma_m(q)$ is the lifetime of a magnetoexciton.

3.1. Wave-vector-dependent light scattering

First, we discuss dipole-forbidden LS induced by the operator $\hat{V}_1$, which can be written as

$$\hat{V}_1=-\frac{\sqrt{N_e}}{\Delta}[f_e\hat{A}_e^+(\mathbf{k}_\parallel)+f_s\hat{A}_s^+(\mathbf{k}_\parallel)]+\text{c.c.} \tag{17}$$

The structure factor of LS (2) is

$$S(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\langle 0|\hat{V}_1^+\hat{V}_1(t)|0\rangle\exp(i\omega_1t)dt$$

$$=-\frac{N_e}{\pi\Delta^2}|L_{01}(k_\parallel)|^2\{|f_e|^2\,\mathrm{Im}[G_e(\omega,k_\parallel)]+|f_s|^2\,\mathrm{Im}[G_s(\omega,k_\parallel)]\}. \tag{18}$$

We see that the structure factor has the Lorentzian form and $\propto k_\parallel^2/\Gamma_m$, when $k_\parallel l_c\ll 1$ and $|\omega-\omega_m(k_\parallel)|\simeq\Gamma_m$. The integrated intensity of dipole-forbidden LS

$$I=\int S(\omega)d\omega\propto k_\parallel^2/B,$$

when $k_\parallel l_c\ll 1$.

3.2. Light scattering in the dipole limit

We now calculate the cross section of LS in the dipole limit $k\to 0$. At zero temperature we can retain in the operator $\hat{W}$ [Eqs. (12) and (13)] only the terms with $\hat{A}_m^+(\mathbf{q})$, $\hat{A}_m^+(-\mathbf{q})\hat{A}_{m'}^+(\mathbf{q})$. Thus, we have

$$\hat{W}_{\rm def}=\frac{\sqrt{N_e}}{\Delta^2}\sum_q u_{\rm eff}(\mathbf{q})\{f_e\hat{A}_e^+(\mathbf{q})+f_s\hat{A}_s^+(\mathbf{q})\},$$

$$\hat{W}_{\rm Coul}=\frac{m_c N_e}{2\mu\Delta^2}\sum_q U(q)\{f_e\hat{A}_e^+(-\mathbf{q})\hat{A}_e^+(\mathbf{q}) + f_s\hat{A}_e^+(-\mathbf{q})\hat{A}_s^+(\mathbf{q})\}. \tag{19}$$

Here we consider the strictly 2D system, where $U_{cc}=U_{cv}=U(q)=2\pi e^2/(\epsilon q)$.

3.3. Light scattering induced by disorder (defects)

The operator $\hat{W}_{\rm def}$ [Eq. (19)] results in LS by magnetoexcitons with arbitrary wave vectors q. This effect can be understood in a single-electron picture. Consider the inelastic LS process, in which $\psi_{c1}(r)(\psi_{c2}(r)$ are the initial (final) single-electron states in the conduction band and $\psi_v(r)$ is the intermediate single-electron state in the valence band. The amplitude of LS is proportional to

$$\langle\psi_{c2}(r)|\psi_v(r)\rangle\langle\psi_v(r)|\psi_{c1}(r)\rangle.$$

The wave functions $\psi_{c1(2)}(\psi_v)$ are solutions of the standard equations

$$\left[\frac{\hat{p}^2}{2m_c}+u_c^{\rm def}(r)\right]\psi_c=\epsilon_c\psi_c,$$

$$\left[-\frac{\hat{p}^2}{2m_v}+u_v^{\rm def}(r)\right]\psi_v=\epsilon_v\psi_v.$$

We see that the wave functions ψ_c and ψ_v coincide if $u_c^{\rm def}m_c=-u_v^{\rm def}m_v$. The latter means $u_{\rm eff}=0$ and $\langle\psi_{c2}(r)|\psi_v(r)\rangle\langle\psi_v(r)|\psi_{c1}(r)\rangle=0$ if $\psi_{c1}(r)\neq\psi_{c2}(r)$. Thus, defect-induced LS arises from the difference between the wave functions of electrons and holes. Light scattering induced by a quasi-classical smooth electric field was considered earlier in Ref. 17. In contrast to Bechstedt *et al.*,[17] we treat electron scattering by defects with a quantum-mechanical approach.

Using the operator $\hat{W}_{\rm def}$ [Eq. (19)], we find the structure factor of defect-induced LS

$$S^{\rm def}(\omega)=-\frac{N_e}{\pi\Delta^4}\int\frac{d^2q}{(2\pi)^2}|L_{01}(q)|^2\langle u_{\rm eff}^2\rangle_q\{|f_e|^2 \times{\rm Im}\,G_e(\omega,q)+|f_s|^2\,{\rm Im}\,G_s(\omega,q)\}, \tag{20}$$

where the correlation function $\langle u_{\rm eff}^2\rangle_q$ is given by

$$\langle u_{\rm eff}^2\rangle_q=|u_{\rm eff}(q)|^2=\int\int d\mathbf{r}d\mathbf{r}'u_{\rm eff}(\mathbf{r})u_{\rm eff}(\mathbf{r}+\mathbf{r}')e^{i\mathbf{q}\mathbf{r}'}. \tag{21}$$

In the backscattering geometry $(\mathbf{k}_{1(2)}\|\mathbf{z})$, the polarized spectra of LS $(\mathbf{e}_1\|\mathbf{e}_2, f_e\neq 0)$ arise from charge-density excitations. The depolarized spectra $(\mathbf{e}_1\perp\mathbf{e}_2, f_s\neq 0)$ relate to spin-density excitations. These selection rules are similar to those for k-dependent LS.[8] We see from Eq. (20) that characteristic wave vectors of magnetoexcitons are determined by random potentials (the function $\langle u_{\rm eff}^2\rangle_q$) and by the magnetic length (the function L_{01}). In the limit $\Gamma_m\to 0$, the structure factor is proportional to the density of states of magnetoexcitons. Near the critical points of dispersions $\omega_{\rm cr}$, where $d\omega_m(q)/dq=0$, the density of states is proportional to $|\omega-\omega_{\rm cr}|^{-1/2}$. The polarized spectrum $S_{\rm pol}^{\rm def}$ in the limit $\Gamma_e\to 0$ has two peaks, which correspond to the excitations with $q_{\rm rot}$ and $q_{\max}$, where $q_{\rm rot}$ and $q_{\max}$ are the wave vectors of the roton minimum and maximum, respectively (see the inset in Fig. 1). In the limit $\Gamma_s\to 0$, the depolarized spectrum $S_{\rm dep}^{\rm def}$ has one singularity which is related to the roton minimum.

Light scattering processes with $\Delta l=2,3,...$ can be described in the same way. The structure factor for the process with $\Delta l=N$ is given by Eq. (20) with a correction $L_{01}\to L_{0N}$.

We now discuss the mechanisms of electron scattering by disorder in quantum wells.

3.3.1. Interface defects. In the case of imperfect interfaces, the quantum-well energies of particles depend on the in-plane coordinate: $W_c=\pi^2/(2m_c^0L^2(\mathbf{r}))$ and $W_v=-\pi^2/(2m_v^0L^2(\mathbf{r}))$, where $L(\mathbf{r})$ is the width of a quantum well, and $m_{c(v)}^0$ are the effective masses in a bulk semiconductor. The quantum-well energies of electrons and holes W_c and W_v play the role of the potentials $u_c^{\rm def}(\mathbf{r})$ and $u_v^{\rm def}(\mathbf{r})$. In this simplest model, we have

$$u_{\rm eff}=\frac{\pi^2}{2L^2(\mathbf{r})}\left(-\frac{1}{m_v^0}+\frac{m_c}{m_c^0m_v}\right).$$

Typically, for GaAs–AlAs quantum wells we can write: $m_v^0\neq m_v$ and $m_c^0\simeq m_c$. Consequently, a reasonable approximation for the effective potential is

$$u_{\rm eff}=\frac{\pi^2}{2L^2(\mathbf{r})}\left(\frac{1}{m_v}-\frac{1}{m_v^0}\right).$$

In the case of GaAs–AlAs quantum wells, we have $m_v^0\simeq 0.4m_0$ and $m_v\simeq 0.17m_0$ for the first heavy-hole subband.[18] For weak fluctuations of the width of a quantum well, we can write

$$u_{\rm eff}=u_{\rm eff}(L_0)+\left(\frac{1}{m_v^0}-\frac{1}{m_v}\right)\frac{\pi^2\delta L(\mathbf{r})}{4L_0^3},$$

where L_0 is the average width of a quantum well, $\delta L(\mathbf{r})=L(\mathbf{r})-L_0$, and $|\delta L|\ll L_0$.

In the limit $|\delta L(\mathbf{r})|\ll L_0$, the correlation function (21) is written as

$$\langle u_{\rm eff}^2\rangle_q=\left[\frac{(1/m_v^0-1/m_v)\pi^2}{4L_0^3}\right]^2\langle\delta L^2(\mathbf{r})\rangle_q. \tag{22}$$

In the case of short-range fluctuations of $\delta L(\mathbf{r})$, we may assume that the characteristic wave vector q_0 of the correlation function $\langle\delta L^2(\mathbf{r})\rangle_q$ is much larger then l_c^{-1} and $\langle\delta L^2(\mathbf{r})\rangle_q\simeq$ const for $q\simeq l_c^{-1}$.

3.3.2. Impurities. Another mechanism of electron scattering is connected with impurities. In this case, the potentials are

$$u_c^{\rm def}(\mathbf{r})=u_v^{\rm def}(\mathbf{r})=\sum_n u_t(\mathbf{r}-\mathbf{R}_n)$$

and

$$u_{\rm eff}=u_c^{\rm def}(1+m_c/m_v),$$

where $\mathbf{R}_n$ are the positions of impurities, and $u_t(\mathbf{r})$ is their potential. The correlation function in the cross section (20) is written as

$$\langle u_{\rm eff}^2\rangle_q=N_t\left(1+\frac{m_c}{m_v}\right)^2|u_t(q)|^2, \tag{23}$$

where N_t is the 2D density of the impurities, and $u_t(q)$ is the Fourier transform of a single-impurity potential. The impurity potential is taken in the form

$$u_t(r)=\frac{e^2}{\epsilon\sqrt{r^2+z_0^2}},$$

where z_0 is the distance between the center of a quantum well and the δ-layer of the impurities.

Figure 1 shows the charge-density-excitation spectrum $S_{\rm pol}^{\rm def}(\omega)$ calculated for various mechanisms of disorder-induced scattering. The lineshape strongly depends on the correlation function $\langle u_{\rm eff}^2\rangle_q$. Curves *1–3* in Fig. 1 represent impurity-induced LS in the systems with $z_0=200$, 300, and 500 Å, respectively. The magnetic length $l_c=100$ Å corresponds to $B=7$ T. The parameter Γ_e can be estimated from the width of the cyclotron peak in high-mobility heterostructures; it is of the order of 0.1 meV. Hence, for $B=7$ T we have $\epsilon_C\simeq 10$ meV and $\Gamma_e/\epsilon_C\simeq 0.01$. In the case $z_0=200$ Å, the main contributions to the spectra arise from the critical points of the dispersion, i.e., from excitations with $q=q_{\max}$ and $q_{\rm rot}$. The spectrum in the case $z_0=500$ Å is strongly shifted to low energies and includes mostly excitations with $q<q_{\max}$. This fact is connected with the exponential function

$$u_t^2(q)=U^2(q)\exp(-2qz_0)$$

in Eq. (20). In the case $z_0=500$ Å, the impurity potential is too smooth to induce roton excitations. In the spectrum for $z_0=300$ Å, the structure is shifted to low energies, but the contribution of rotons is still essential. Curve *4* in Fig. 1 shows the spectrum of LS induced by a short-range random potential ($\langle\delta L^2(\mathbf{r})\rangle_q=$ const). We see that the main contributions to the spectrum *4* in Fig. 1 are related to the critical points of the dispersion. In Fig. 1 the peak intensities strongly depend on the parameter Γ_e/ϵ_C, while the lineshape of the spectra away from the peaks is relatively insensitive to this parameter.

In their experiment Pinczuk *et al.*[2] observed a broad structure at energies above the cyclotron frequency, which was interpreted as the roton density of states. For the multiple quantum wells studied by them the distance z_0 was about 300 Å. Our calculations show that for the distance $z_0=300$ Å the spectrum is shifted to energies below the roton energy. It can be assumed, therefore, that in the experiment[2] the LS spectrum arises, in part, from a short-range random potential (interface defects). In addition, the experimental spectra depend essentially on the laser frequency, which is the signature of strong interband resonance. Our theoretical results are valid away from strong resonance and, consequently, a detailed comparison between theoretical and experimental spectra is not possible. Our theory, nevertheless, makes it possible to estimate the characteristic wave vectors of excitations in Raman spectra and to understand the mechanism of LS.

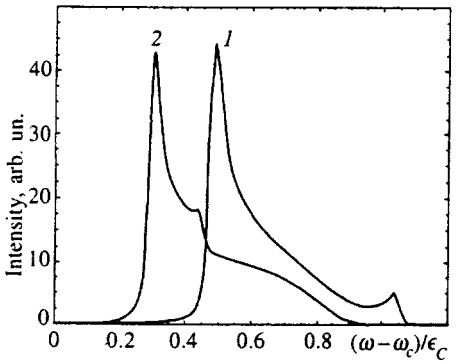

FIG. 2. The spectrum of light scattering by two magnetoexcitons: curves *1* and *2* show the excitations $2\omega_e$ and $\omega_e+\omega_s$, respectively; the parameter $\Gamma/\epsilon_C=0.01$.

3.4. Light scattering by two elementary excitations

Light scattering induced by the Coulomb interaction is connected with the operator $\hat{W}_{\rm Coul}$ [see Eq. (19)]. This operator results in LS by two charge-density excitations (the polarized spectrum) and by combined excitations $\omega=\omega_s+\omega_e$ (the depolarized spectrum). Light scattering by two spin-density excitations is absent here, because the operator V_2 contains the first power of the spin operator $\hat{\sigma}_z$. We assume that the next terms in the expansion (6) can lead to LS by two spin-density excitations.

The structure factors for polarized and depolarized spectra are

$$S_{\rm pol(dep)}^{\rm Coul}(\omega)=-|f_{e(s)}|^2\frac{N_e^2}{\pi\Delta^4}\left(\frac{m_c}{2\mu}\right)^2\int\frac{d^2q}{(2\pi)^2}|L_{01}(q)|^4 \times U^2(q)\,{\rm Im}[G_{ee(es)}(\omega,q)], \tag{24}$$

where $G_{ee(es)}(\omega,q)=1/(\omega-\omega_e-\omega_{e(s)}+2i\Gamma)$ are the two-magnetoexciton Green's functions. For simplicity, we assume $\Gamma_e=\Gamma_s=\Gamma$. In the limit $\Gamma\to 0$, the structure factors (24) are proportional to the density of states of two magnetoexcitons. We see from Eq. 24 that the characteristic wave vectors of magnetoexcitons in LS spectra are of the order of l_c^{-1}. In the limit $\Gamma\to 0$, the functions $S_{\rm pol(dep)}^{\rm Coul}$ have peaks at critical-point energies (Fig. 2).

The matrix elements $\hat{W}_{\rm Coul}$ [Eqs. (13) and (19)] originate from Coulomb correlations. Such a process can be considered as a «shakeup». In other words, interband optical transitions are accompanied by shakeup of an electron system with emission of elementary excitations.

Light scattering by low frequency excitations in the regime of the fractional QHE was reported in Refs. 3 and 4. The spectra observed in Ref. 4 were interpreted in terms of

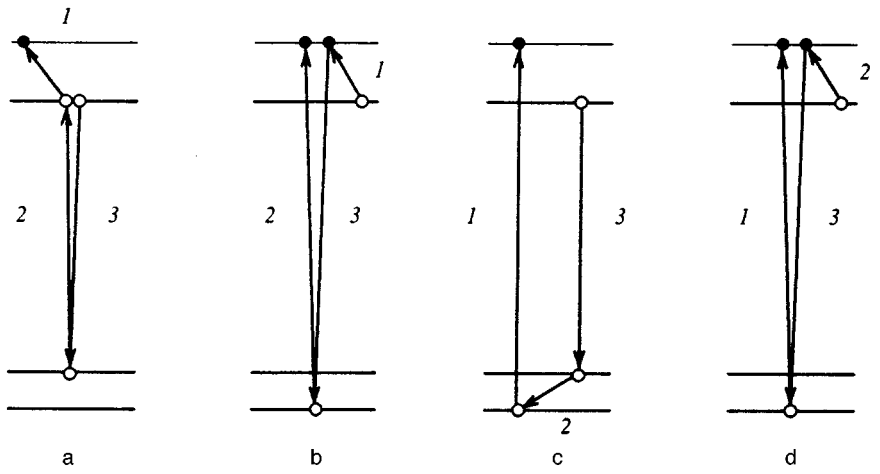

FIG. 3. Diagrams for light scattering induced by defects, in which the interband transitions are optical and the intraband transitions are assisted by defects. Scattering by defects occurs in the initial electron states (diagrams a and b) and in the intermediate states of the light scattering process (diagrams c and d). For the case $\nu=2$, the contributions connected with defect-induced scattering in the final states in third-order perturbation theory are absent.

two-roton excitations, which have low energies (about $0.2\epsilon_C$.) The roton excitations in the fractional QHE are connected with intra-Landau-level transitions. Until now, we do not know of any publications devoted to studies of the structure near the frequency $2\omega_c$. In an experimental situation, the Raman spectrum near the frequency $\omega=2\omega_c$ can consist of two contributions: the first can be the defect-induced structure related to the magnetoexciton $\Delta l=2$ and the second can be the contribution of the two-magnetoexciton process. These contributions can be separated because the critical-point energies of the two types of excitations are different.

Here we do not consider the spin-flip inter-Landau-level spectra, which occur in our approach if the geometry deviates from backscattering and the incident light is in resonance with the light-hole valence band.[10] A theoretical description of these processes is similar to that for the case of charge-density (spin-density) excitations.

4. RESONANT STRUCTURE OF THE CROSS SECTION

In Sec. 2, we consider the operator $\hat{V}_{\rm eff}$ in the limit $|\Delta|=|E_g-\omega_1|\gg\omega_c$. We now discuss the case of strong resonance between Landau levels in the conduction and valence bands when $|\Delta_{ll'}|\sim\omega_c(\omega_{hh})$, where $\Delta_{ll'}=\omega_1-E_g-\omega_c(1/2+l)-\omega_{hh}(1/2+l')$, and ω_{hh} is the cyclotron frequency of heavy holes. The energies $E_{gll'}=E_g+\omega_c(1/2+l)+\omega_{hh}(1/2+l')$ correspond to the interband optical resonances. The effective g-factors are neglected. At the same time, we assume that $|\Delta_{ll'}|\gg\epsilon_C\gg\Gamma$ and $\omega_c(\omega_{hh})\gg\epsilon_C$. Thus, the expansion parameters are $\epsilon_C/\Delta_{ll'}$, ϵ_C/ω_c, and Γ/ω_c.

The resonant contributions to the amplitude in third-order perturbation theory are shown in Fig. 3. In these diagrams, the interband virtual transitions are optical, while the intraband transitions are assisted by a random potential. Using these diagrams, we rewrite the operator $\hat{W}_{\rm def}$ [Eq. (12)] with the substitution

$$\frac{u_{\rm eff}(q)}{\Delta^2}\to\left(\frac{1+m_c/m_v}{\Delta_{10}\Delta_{21}}-\frac{1}{\Delta_{11}\Delta_{21}}\right)u_c^{\rm def}(q)+\frac{u_v^{\rm def}(q)}{\Delta_{11}\Delta_{10}}.$$

The diagrams for two-magnetoexciton processes are shown in Fig. 4. The correction for the operator $\hat{W}_{\rm Coul}$ [Eq. (13)] is

$$\frac{1}{\Delta^2}\left[\left(\frac{m_c}{2\mu}-1\right)U_{cc}+U_{cv}\right]\to\left(\frac{m_c}{2\mu}\,\frac{1}{\Delta_{20}\Delta_{31}}-\frac{1}{\Delta_{11}\Delta_{31}}\right)U_{cc}+\frac{1}{\Delta_{11}\Delta_{20}}\,U_{cv},$$

which shows a fine structure of interband resonances.

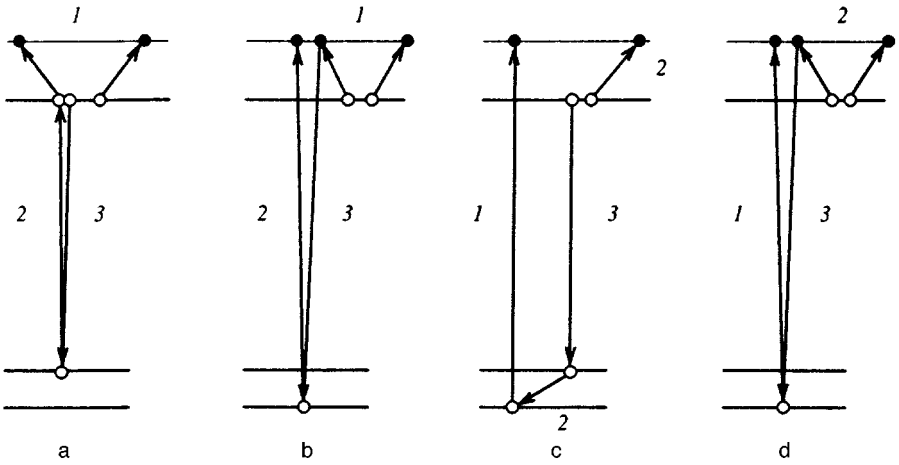

FIG. 4. Diagrams for light scattering induced by the Coulomb interaction, in which the interband transitions are optical and the intraband transitions are assisted by the Coulomb coupling. In the diagrams a and b, the Coulomb interaction induces virtual transitions in the initial states. The diagrams c and d include similar transitions in the intermediate states.

5. DISCUSSION

In this paper, we have shown that resonant LS from an electron plasma in the dipole limit can be described by effective operators of two types [see Eqs. (12) and (13)]:

$$\hat{W}_{\rm def}=\frac{1}{\Delta^2}\sum_{\mathbf{q}} u_{\rm eff}(\mathbf{q})[f_e\hat{\rho}_e(\mathbf{q})+f_s\hat{\rho}_s(\mathbf{q})],$$

$$\hat{W}_{\rm Coul}=\frac{1}{2\Delta^2}\sum_{\mathbf{q}} U(q)[f_e\hat{\rho}_e(-\mathbf{q})\hat{\rho}_e(\mathbf{q}) + f_s\hat{\rho}_e(-\mathbf{q})\hat{\rho}_s(\mathbf{q})], \tag{25}$$

where $\hat{\rho}_e(\mathbf{q})$ and $\hat{\rho}_s(\mathbf{q})$ are the Fourier transforms of the charge and spin densities, respectively; the functions $f_{e(s)}$ determine the selection rules of LS. The operators (25) are the leading terms in the expansion of the LS amplitude in terms of the parameter $1/\Delta$.

The operator $\hat{W}_{\rm def}$ describes LS processes in the presence of disorder (defects). The matrix element $u_{\rm eff}$ is a combination of Fourier transforms of random potentials in the conduction and valence bands [Eq. (14)]:

$$u_{\rm eff}(\mathbf{q})=u_v^{\rm def}(\mathbf{q})+\frac{m_c}{m_v}u_c^{\rm def}(\mathbf{q}).$$

The cross section of disorder-induced LS is given by Eq. (1) with $f(q)\propto|u_{\rm eff}(\mathbf{q})|^2$. The characteristic wave vectors of excitations in these spectra are connected with the correlation functions of disorder.

The operator $\hat{W}_{\rm Coul}$ relates to LS induced by the Coulomb interaction. This operator leads to two-quasiparticle spectra in an ideal system. The characteristic wave vectors of excitations in LS spectra in this case are determined by the electron-electron interaction potential and in a high magnetic field are of the order of $1/l_c$.

The results of this paper can be used for a description of light scattering in semiconductor plasmas and in laterally modulated electron systems (quantum wires and dots). Our approach is valid when $E_g\gg|\Delta|=|E_g-\omega_l|\gg\epsilon_{\rm exc}$, where $\epsilon_{\rm exc}$ is the characteristic energy of electron excitations in the LS process.

The author would like to thank V. A. Volkov for motivating remarks and E. L. Ivchenko, A. Pinczuk, and A. Maslov for helpful discussions. This work was supported by the A. von Humboldt Stiftung and by the Russian Fund for Fundamental Research.

[1] A. Pinczuk, B. S. Dennis, L. N. Pfeiffer, and K. W. West, Semicond. Sci. Technol. **9**, 1865 (1994).
[2] A. Pinczuk, J. P. Valladares, D. Heiman *et al.*, Phys. Rev. Lett. **61**, 2701 (1988).
[3] A. Pinczuk, B. S. Dennis, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **70**, 3983 (1993).
[4] A. Pinczuk, B. S. Dennis, Song He, L. L. Sohn, L. N. Pfeiffer, and K. W. West, to be published.
[5] I. K. Marmorkos and S. Das Sarma, Phys. Rev. B **45**, 13396 (1992).
[6] P. M. Platzman and Song He, Phys. Rev. B **49**, 13674 (1994).
[7] A. O. Govorov, Fiz. Tverd. Tela **38**, 2673 (1996) [Phys. Solid State **38**, 1466 (1996)].
[8] D. C. Hamilton and A. L. McWhorter, in *Light Scattering Spectra of Solids*, ed. by G. B. Wright, Springer, New York (1969), p. 309.
[9] A. O. Govorov, in *Workbook of the Int. Symp.: Heterostructures in Science and Technology*, Würzburg, Germany (1995), p. 56.
[10] A. O. Govorov and A. V. Chaplik, Solid State Commun. **85**, 827 (1993).
[11] M. V. Klein, in *Light Scattering in Solids*, ed. by M. Cardona, Springer-Verlag, Berlin-Heidelberg (1975), p. 147.
[12] V. A. Voitenko and I. P. Ipatova, [Sov. Phys. JETP **70**, 125 (1990)].
[13] J. K. Jain and P. B. Allen, Phys. Rev. B **32**, 997 (1985).
[14] P. A. Wolff, Phys. Rev. **171**, 436 (1968).
[15] Yu. A. Bychkov, S. V. Iordanskii, and G. M. Eliashberg, JETP Lett. **33**, 143 (1981); S. V. Iordanskii and B. A. Muzykantskii, J. Phys.: Condens. Matter **3**, 9103 (1991).
[16] C. Kallin and B. I. Halperin, Phys. Rev. B **30**, 5655 (1984); Phys. Rev. B **31**, 3635 (1985).
[17] F. Bechstedt, R. Enderlein, and K. Peuker, Phys. Status Solidi B **78**, 711 (1976).
[18] A. M. Cohen and G. E. Marques, Phys. Rev. B **41**, 10608 (1990).

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

Anisotropic pinning in macroscopic electrodynamics of superconductors

E. Yu. Klimenko, S. V. Shavkin, and P. V. Volkov

Kurchatov Institute, 123182 Moscow, Russia
(Submitted 14 January 1997)
Zh. Éksp. Teor. Fiz. **112**, 1055–1081 (September 1997)

Anisotropy of critical currents and electric fields in superconductors with strong pinning has been ascribed in the macroscopic model to features of the material equation system relating the electric field to the current density in a superconductor. The anisotropy of the pinning proper is described by an operator relating the pinning force density to the vectors of magnetic induction and Lorentz force. In the approximation of an extended critical state model, a feasible expression of this operator is given in the form of an algorithm based on the concept of a collective anisotropic potential well containing fluxoids. The current-carrying capacity of a strongly anisotropic niobium–titanium foil as a function of the orientation of the current density and applied field with respect to the principal axes of the material has been investigated in detail. Given measurements of the transverse electric fields in the foil under magnetic fields normal to the foil plane, we can plot cross sections of surfaces describing the pinning force density in the space of magnetic induction and Lorentz force. [S1063-7761(97)01909-4]

1. INTRODUCTION

The relation between electric field and current density in high-pinning superconductors is not trivial and has not been fully investigated to this time. Therefore, even elementary electrodynamic problems of dimensionality higher than one cause considerable difficulties, starting with definition of a problem, and usually lead to introduction of additional hypotheses, such as the hypothesis of a zero-force configuration of magnetic induction and current density in superconductors under magnetic field parallel to the current[1] or the hypothesis of the fluxoid cutting.[2] However, such features of real superconductors as self-consistency of the current density and magnetic induction distributions in the superconductor cross section, smoothness of the superconducting transition, and anisotropy of real high-pinning superconductors are rarely taken into account, so agreement between calculations based on such hypotheses and experimental data is, at best, qualitative.

Recent interest in the anisotropy of a superconductor's critical current density is related to studies of high-T_c superconductors, whose structures are very anisotropic. A lot of interesting effects,[3,4] including the anisotropy of the critical field and pinning on basal planes, result from peculiar features of HTSC structures. In combination with the pinning on inserted inhomogeneities, which is a traditional cause of the anisotropy of the critical current, these features of HTSC notably complicate the analysis of their properties. It is natural that the researchers prefer to perform quantitative analysis of pinning on inhomogeneities, which is a difficult problem in itself, in the absence of additional complications. Strong anisotropy of critical currents is manifested in wires and foils made from traditional superconductors with an almost isotropic critical field, in particular, niobium–zirconium, niobium–titanium, and other alloys.[5–10] Information about the small anisotropy of the upper critical field in niobium–titanium and niobium–zirconium wires can be found in Refs. 11 and 12. We assume that in these materials the pinning anisotropy is observed in its purest form, and experiment can yield data which can be conveniently compared to model calculations.

It is known that the current-carrying capacity of superconductors is controlled by the critical Lorentz force,[13] but not by the critical current itself. In traditional problems concerning electrodynamics of high-current superconductors, when mutually orthogonal fields and currents were considered, the difference between approaches based on the critical current and critical Lorentz force is not important, and this difference is important only when the angle between the field and current density has an arbitrary value. The available models of anisotropic critical current,[14–16] which yield a fairly accurate description of the HTSC magnetization, do not provide accurate results concerning the current-carrying capacity of a superconductor when the current density is arbitrarily tilted with respect to the magnetic field. The description of anisotropic HTSC properties[17] in terms of the phenomenological scaling approach based in the Ginzburg–Landau equations applies only to the case of axial symmetry in layered oxide superconductors, describes the internal pinning in HTSC, and does not hold in the case of hard superconductors, where the pinning anisotropy has a different nature.

Given the variety of anisotropic patterns of electric fields and current densities in superconducting samples depending on their orientation in external magnetic field, one must distinguish between external and intrinsic causes. The intrinsic causes of anisotropy are the configuration and distribution of defects acting as pinning centers. Among the external causes, which appear even in materials with isotropic pinning, we can mention small changes (about several percent) in the critical current of plane samples due to changes in the magnetic field orientation with respect to the sample plane[18] caused by a variation in the current distribution in the plane

due to the field generated by the current when the function $j(B)$ is steep.

In studies of anisotropic pinning, the importance of the real current distribution in the sample plane is significantly greater. The set of causes of anisotropy that are external with respect to the array of pinning centers of the material also includes the possibility of forming so-called zero-force configurations of current and field in a superconducting wire, which has been studied for more than thirty years, starting with the work of Bergeron.[1]

The real anisotropy of the material can be distorted by the effect of the field orientation on the development of instability in a superconductor, especially in tests of nonstabilized superconducting materials with high current-carrying capacity; therefore, only experiments in which the steady-state current–voltage characteristic is measured yield accurate results.

Experimental manifestations of intrinsic anisotropy are varied. For example, a giant anisotropy of the critical current was detected in a niobium–zirconium foil[6] when the Lorentz force and induction vector tilts with respect to the foil plane were varied, while the magnetic field remained perpendicular to the current density. A notable anisotropy of the critical current density was observed in studies of magnetization of niobium–titanium wires in both longitudinal and transverse magnetic fields.[8] In these experiments, the Lorentz force remained perpendicular to the wire axis, the current density and induction were mutually orthogonal, but the induction vector tilt in the material varied. To complete our list, we must mention the manifestations of various mechanisms of a current transition when the sample is reoriented in an external field,[19] although the results of the experiments could be described without such a far-fetched hypothesis.

In this paper, we present a part of a detailed investigation of the current-carrying capacity of a thin niobium–titanium foil as a function of magnetic field orientation in conditions when all external causes of anisotropy were either eliminated or properly taken into account, and the effects of the orientation of both the Lorentz force and induction vector on the critical current are clearly distinguished.

We have attempted here to systematically apply the traditional approach to the electrodynamics of high-pinning superconductors in arbitrarily oriented fields, which has long been used successfully in the special case of mutually orthogonal induction and current density, both in the form of a model of a critical state[20] and a model of the actual transient characteristics.[21] Here we describe a slightly simplified model in the limits of a modified critical state model, and using the concept of the critical current. We suggest a phenomenological model with several independent empirical parameters for a quantitative description of an anisotropic array of pinning centers. When the concept of competition between the Lorentz force and anisotropic pinning is applied, one can calculate the current-carrying capacity of a superconducting material, and self-consistent distributions of current density and magnetic induction in the sample cross section for arbitrary orientations of applied magnetic field and current density.

Naturally, the macroscopic approach describes interaction between magnetic flux and matter not in terms of an individual fluxoid, but in terms of a unit volume of the material, and the interaction energy and pinning force per unit volume.

On the assumption that the only cause of the type II superconductor transition to the resistive state with an increase in the transport current is the competition between the Lorentz force and the pinning force, the relation between the electric field and current density in the steady state can be expressed by the following equation system:

$$\mathbf{F}_L = \mathbf{j} \times \mathbf{B}, \tag{1}$$

$$\eta \mathbf{v} = \theta(F_L - F_P)\left(1 - \frac{F_P}{F_L}\right)\mathbf{F}_L, \tag{2}$$

$$F_P = \Xi(\mathbf{F}_L, \mathbf{B}), \tag{3}$$

$$\mathbf{E} = \mathbf{B} \times \mathbf{v}, \tag{4}$$

where $\mathbf{E}$ is the electric field, $\mathbf{j}$ is the current density, $\mathbf{B}$ is the magnetic induction, $\mathbf{v}$ is the drift velocity of the magnetic flux, $\eta = B^2/\rho_{\text{eff}}$ is the viscosity, $\mathbf{F}_L$ is the Lorentz force on the flux, $\theta(F_L - F_P)$ is the Heaviside step function, which is zero for $F_L < F_P$ and unity for $F_L > F_P$, and Ξ is an operator describing the effect of the Lorentz force direction and the magnetic induction direction and magnitude on the pinning force.

In Eq. (3) of this system, the nonlinear operator Ξ describes the real anisotropy of pinning in a superconductor, if the superconductor is isotropic in all other respects. Equations (1), (2), and (4) can be applied to angular dependences of critical currents and electric fields observed in superconductors with both isotropic and anisotropic pinning.

As concerns the constitutive equation relating the applied magnetic field to the induction, we assume that $\mathbf{B} = \mu_0 \mathbf{H}$, bearing in mind that for $B \leqslant \mu_0 H_{c1}$ this assumption may lead to significant errors.

Below we consider some manifestations of orientation effects in isotropic superconductors as an illustration of the macroscopic approach, and also consider one form of the operator Ξ derived from a detailed experimental study and computations of critical currents in a niobium–titanium foil with high anisotropic pinning.

2. ORIENTATION EFFECTS IN SUPERCONDUCTORS WITH HIGH ISOTROPIC PINNING

Although we know of no isotropic superconductors with high pinning, we assume that it is worthwhile to separate effects due to the dynamics of magnetic flux in a superconductor and due to the inherent anisotropy of the material.

2.1. Transverse electric field

In a superconductor with isotropic pinning, the pinning force F_P in Eq. (2) is independent of both the directions of the Lorentz force and magnetic induction and can be interpreted geometrically as a sphere in the force space such that a Lorentz force vector piercing the sphere transfers a superconductor to the resistive state. It is clear that the electric field in Eq. (4) is aligned with the current density only if the induction is perpendicular to the current density:

$$\mathbf{E}=\mathbf{B}\times\mathbf{v}=\frac{1}{\eta}\theta(F_L-F_P)\left(1-\frac{F_P}{F_L}\right)(\mathbf{j}B^2-\mathbf{B}(\mathbf{j}\cdot\mathbf{B})). \quad (5)$$

Since the electric field is always perpendicular to the magnetic induction, the angle between the electric field and current density is $\pi/2-\alpha$, where α is the angle between the current density and induction, i.e., a transverse electric field is generated in the resistive state. The physical basis for this angle is the hard coupling between components of a moving magnetic flux carried through the superconductor by fluxoids. In a normal metal, such an effect cannot be observed, because in this case different components of magnetic flux can propagate independently. This effect should not be confused with the well known effect of guided vortex motion[22] due to the pinning anisotropy, which is detected even in the case of mutually perpendicular magnetic field and current density. Note that the angle between the electric field and current can be tens of degrees, which is considerably larger than the typical angle for the Hall effect in type II superconductors.

2.2. Critical current in a superconductor for arbitrary orientation of external field

Consider a plate of thickness $2d$ $(-d<y<d)$ in the xz plane containing a current in the x-direction in an external field of an arbitrary orientation with components H_{xe}, H_{ye}, and H_{ze}. Let current J_x be critical, i.e., occupy the entire volume of the plate. Equation (3) in this case has the form of a critical state equation $F_L=F_P(B)$, which can be expressed as follows:

$$\frac{\partial B_z}{\partial y}\sqrt{B_y^2+B_z^2}=\mu_0 F_P(\sqrt{B_y^2+B_z^2+B_z^2(y)}), \quad (6)$$

whose boundary conditions are $B_z(d)=\mu_0(H_{ze}+J_c/2)$ and $B_z(-d)=\mu_0(H_{ze}-J_c/2)$, while $B_x=\mu_0 H_{xe}$ and $B_y=\mu_0 H_{ye}$. This equation can be solved, and the result is determined by the specific form of the function $F_P(B)$.

Figure 1 shows the calculations of the critical current density as a function of magnetic field for various orientations of the external magnetic field (the pinning force as a function of magnetic induction is given by the expression[23] $F_P(B)=AH_{c2}^2\sqrt{b}(1-b)$, where $b=B/B_{c2}$). The dome-shaped curve of the critical current versus parallel magnetic field, which in due time brought about the hypothesis of zero-force charge flow in superconductors, reflects the dome-shaped pinning force dependence on magnetic field. The large critical current in a longitudinal magnetic field corresponding to $b\sim 0.33$ is due to the maximum of the pinning force in this region, whereas the transverse field component generated by the current and determining the Lorentz force is very small. Note that the described solutions are purely formal when B_z has opposite signs on different plate surfaces, since Eq. (6) does not take account of formation of almost force-free current and magnetic induction configurations corresponding to negligible forces, nor of a paramagnetic moment generated when Lorentz forces on either side of the plate are directed into the plate.

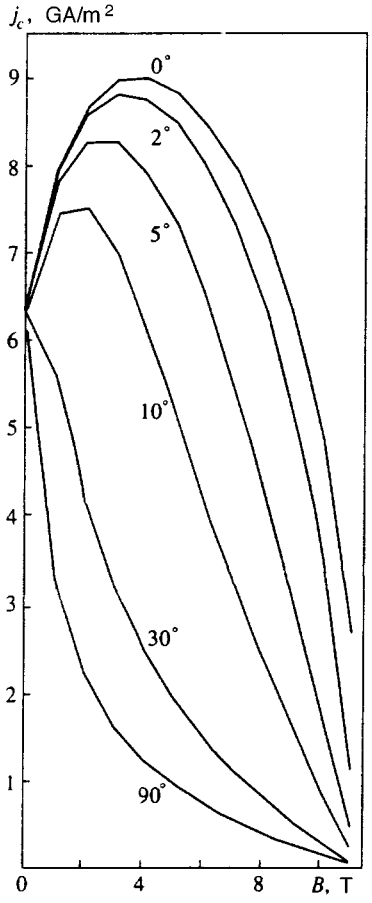

FIG. 1. Critical current density in an isotropic superconducting plate with high pinning at different angles between current density and external magnetic field parallel to the plate surface. The plate thickness is 1 mm, $AH_{c2}^2=1.44\times10^9$ A/m^2.

In all calculations given in this paper, we have assumed that local current densities in a superconductor carrying critical current are always aligned with the transport current.

Owing to the high sensitivity of parameters of the constitutive equation to the magnetic induction, the field generated by the transport current has considerable effect on the current distribution in the superconductor cross section, especially in longitudinal or small external fields. Below we discuss an example with a nonuniform current density distribution in an anisotropic material generating a maximum in the total current through a sample as a function of magnetic field in low external fields.

3. PINNING ANISOTROPY

A self-consistent phenomenological model of a superconductor with anisotropic pinning should contain a minimum number of empirical parameters needed for a satisfactory description of the effect, and an algorithm for calculating the current-carrying capacity of a material in a magnetic field of an arbitrary orientation. In the most general case, the current-carrying capacity is defined in terms of the parameters of current–voltage characteristics (CVC). Most authors, however, have described the current-carrying capacity of a superconductor in terms of one parameter, namely

the critical current, which is defined as a current corresponding to a certain point on CVC determined by an arbitrarily selected value of electric field. In this paper, the model is developed in this approximation.

3.1. Model of anisotropic pinning

Since pinning corresponding to the critical current can be described in an isotropic material for a fixed magnetic induction in terms of a sphere in the force space whose radius is the critical Lorentz force, it is natural to visualize pinning in an anisotropic material using a surface of lower symmetry in the force space. The shape and dimensions of the surface are determined by the distribution of pinning centers and their properties. In the general case, the surface shape is arbitrary, but it should be closed since critical currents are finite, and probably symmetric about the origin, because the critical current does not change when its direction is reversed. A set of such surfaces for different values of magnetic induction ranging between zero and B_{c2} can be defined as a pinning body. For each magnetic induction corresponding to a certain pinning surface, one can introduce a collective potential well for a sufficiently large array of fluxoids. In the absence of transport current, this array occupies the position with the lowest energy at the bottom of this well, and a minimum of the magnetic flux energy density does not mean that each fluxoid trapped by individual pinning centers is at the bottom of its potential well. When a transport current is on, the array is driven up along the well slope in the direction of the Lorentz force. A quasi-harmonic elastic response of a pinned fluxoid lattice corresponding to this stage was detected in experiments with both alternating current[24] and pulsed field.[25] If the Lorentz force is larger than the maximum slope of the potential well, the entire array is set in motion, which leads to initiation of an electric field and energy dissipation (see Eqs. (1) and (2)).

The force of anisotropic pinning measured in experiments, however, depends not only on the Lorentz force direction in a superconductor, but also on the magnetic induction direction, which leads to a considerable complication of the suggested simple model. Each direction of the Lorentz force can be characterized by a cross section of the pinning surface, which is a plane closed curve and determines the critical value of this force as a function of the magnetic induction orientation in the plane perpendicular to the Lorentz force.

In this connection, the geometrical model of the pinning body should be also considered in the space of twice the dimensionality. Let us illustrate this statement by a simple model. We define the maximum pinning force opposing the Lorentz force in direction $\mathbf{l}$ in the form

$$F_P = -\max\left(\frac{\partial U}{\partial \mathbf{l}}\right) = -\mathbf{e}(\mathbf{l})\frac{U_0(\mathbf{B})}{L(\mathbf{l},B)}, \tag{7}$$

where $L(\mathbf{l},B)=U_0(\mathbf{B})/|\max(\partial U/\partial l)|$ is the effective width of the collective potential well in the Lorentz force direction, $\mathbf{e}(\mathbf{l})$ is the unit vector in the Lorentz force direction, and $U_0(\mathbf{B})$ is the potential well depth. An important assumption of the model is that the collective potential well depth corresponding to pinning at a given point of the superconductor cross section depends only on the magnitude and direction of the magnetic induction (and does not explicitly depend on the current density and Lorentz force), whereas the effective width of the collective potential well is assumed to be dependent on the induction magnitude (density of vortices) and the Lorentz force direction, and independent of the direction of the magnetic induction. These assumptions are based on the following reasoning.

It is known[26,27] that even weak pinning centers destroy long-range order in a regular fluxoid lattice (this is the reason why we prefer the term ''array of fluxoids'' to ''fluxoid lattice''). In the case of high pinning, its contribution to the free energy is relatively large. Estimates for a niobium–titanium wire demonstrate[28] that short-range order takes place only over a range of several lattice periods, i.e., different fragments of fluxoids are pinned almost independently, and the fluxoid array is shaped to conform to the ensemble of pinning centers. The specific coupling energy U of these two ensembles depends on individual characteristics of the pinning centers, their density, and distribution throughout the superconductor, and anisotropy of any of these characteristics should lead to a dependence of the fluxoid array coupling energy on the magnetic induction direction in the material.

The effective dimension of the collective potential well is no larger than half the minimum separation between two energetically equivalent positions of the fluxoid ensemble. A rough estimate of this distance is the lesser of two values, namely the mean distance between neighboring fluxoids or the mean distance between neighboring (in the direction of possible drift of fluxoids) pinning centers. Since the mean separation between fluxoids is a function of magnetic induction, and the distance between pinning centers is constant, in a material with sufficiently sparse pinning centers, the well width can be constant in small fields and decrease with the fluxoid density at higher fields.

Some authors[24,25,29] determined elastic displacement of the fluxoid lattice in various materials (assuming that pinning is isotropic), and the maximum elastic displacement proved to be several tenths of the lattice constant. Naturally, in the case of anisotropic pinning, the critical displacement should also be an anisotropic parameter. Since elastic displacement is due to the Lorentz force, the pinning force should be determined by the width of the generalized potential well in none other than the direction of the Lorentz force.

In terms of the proposed phenomenological model, all features of pinning in a specific material are determined by specific angular dependences of the well depth U_0 and its width L. Thus, in order to obtain exhaustive information about the current-carrying capacity of a material in an arbitrary magnetic field, one must construct with the highest possible accuracy the angular and field dependences of these well parameters. It is clear that this problem might demand a huge experimental effort. However, if the analysis is limited to a few reasonable assumptions about feasible shapes of the ''energy'' surface (U_0) and ''width'' surface (L), the current-carrying capacity can be described with a reasonable accuracy using a few parameters. One feasible assumption is that extreme values of U_0 and L are achieved when the mag-

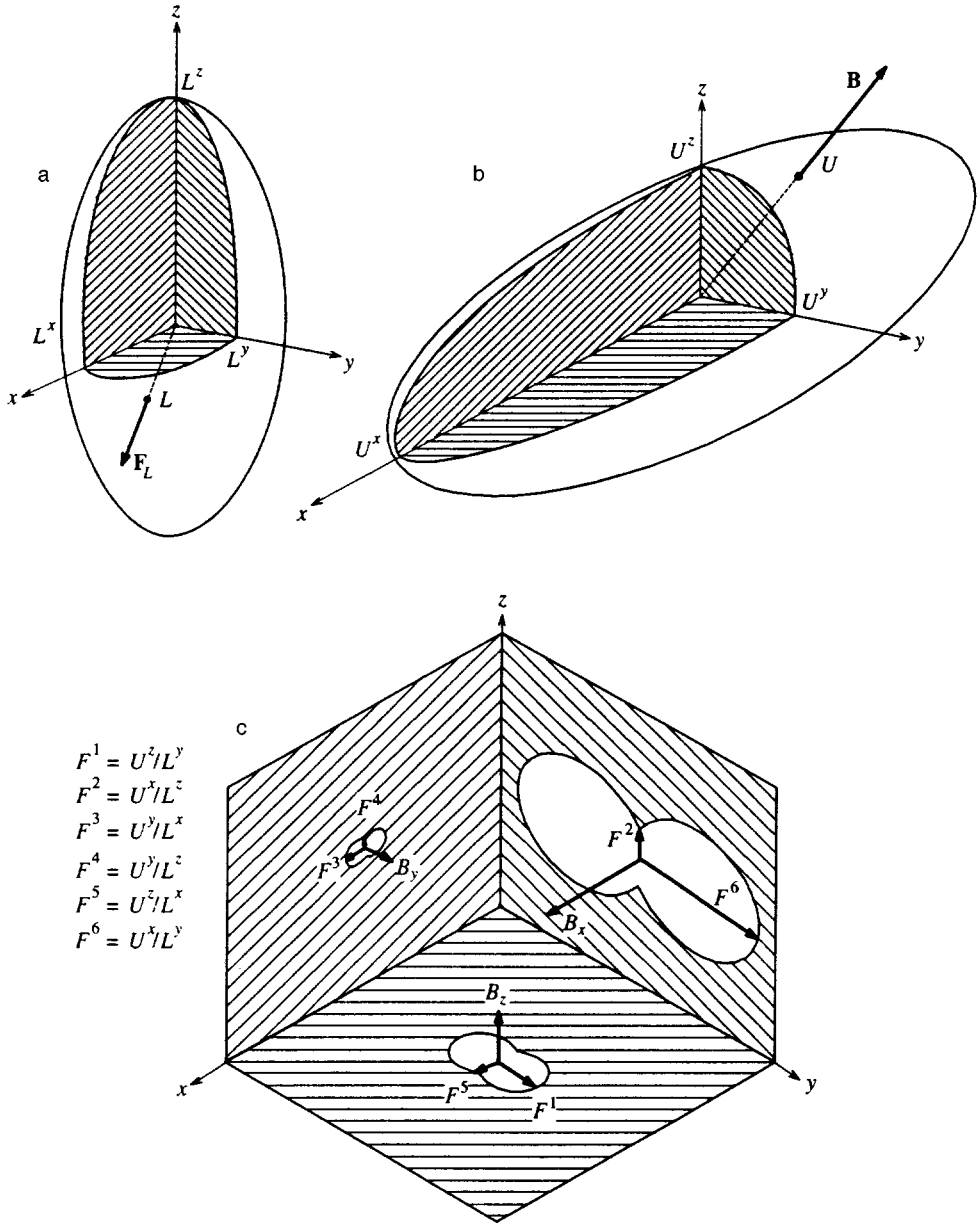

FIG. 2. Illustration of the definition of the current-carrying capacity of an anisotropic superconductor with a multidimensional pinning surface: (a) surface describing the characteristic width of the collective pinning potential well as a function of the Lorentz force direction; (b) surface describing the energy of coupling between a fluxoid array and pinning centers as a function of the magnetic induction direction; (c) cross sections of the pinning body by planes perpendicular to the induction vectors aligned with the principal axes of the material (definition of the six principal values of the pinning force).

netic induction or Lorentz force are aligned with specific orthogonal directions in the material (we call these the principal axes), while for intermediate directions the depth and width of the potential well gradually vary between their maxima and minima. It is natural to assume that in a cold-rolled foil the principal axes are the roll direction (i.e., the x-axis), the normal to the foil surface (i.e., y-axis), and the perpendicular to the roll direction in the foil plane (i.e., z-axis). The potential well depth is U^x, U^y, or U^z when the induction is directed along the x, y, or z axis, respectively. When the direction of the induction vector changes gradually, one can hardly expect abrupt changes in the coupling energy, and it should be described by a smooth function of the angles. It is possible that the determination of the shape of such a surface will at some point become a goal in such studies of superconducting materials, but in this paper, in order to make the results of our analysis easily interpretable, we consider the simplest model surface, namely an ellipsoid defined by

$$\left(\frac{\cos\alpha}{U^x}\right)^2+\left(\frac{\cos\beta}{U^y}\right)^2+\left(\frac{\cos\gamma}{U^z}\right)^2=\frac{1}{U_0^2}, \tag{8}$$

where $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ are the direction cosines of the magnetic induction. Thus, in the absence of external current, the fluxoid array binding energy U_0 per unit volume of a superconductor is proportional to the length of the radius vector of the surface (8) aligned with the magnetic induction (Fig. 2b). Similarly we define in the 3D space a surface of well widths with principal axes L^x, L^y, and L^z. We assume that for an arbitrary direction of the Lorentz force, the corresponding width L of the potential well can also be described by an ellipsoidal surface (Fig. 2a):

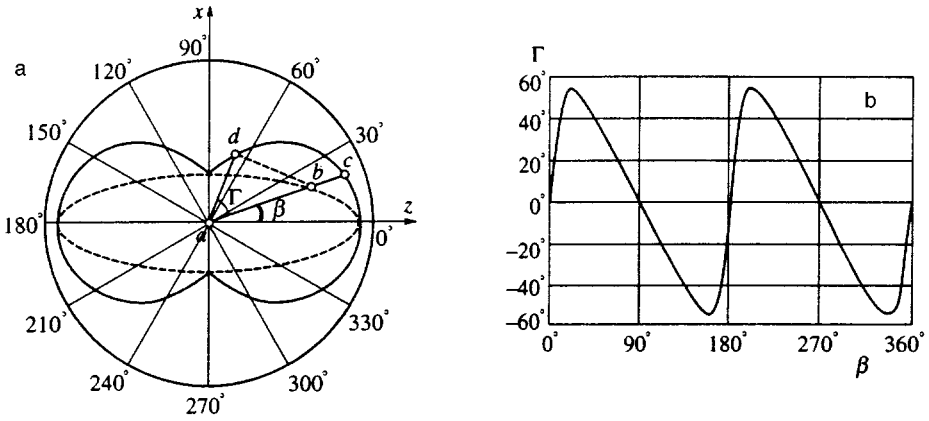

FIG. 3. (a) Diagram illustrating generation of transverse electric field for some current orientations in a superconductor with anisotropic pinning. A cross section of the pinning surface by the xz-plane is shown (β is the angle between the Lorentz force and axis perpendicular to the roll direction, Γ is the angle between the current of magnetic flux and Lorentz force, a is the origin, c is the pinning force value in the direction of the Lorentz force, b is the maximum allowed Lorentz force, d is the pinning force value in the direction of the minimal Lorentz force projection). (b) Calculated angle between the flux motion direction and Lorentz force as a function of the angle between the Lorentz force and the axis perpendicular to the roll direction.

$$\left(\frac{\cos\alpha'}{L^x}\right)^2+\left(\frac{\cos\beta'}{L^y}\right)^2+\left(\frac{\cos\gamma'}{L^z}\right)^2=\frac{1}{L^2}, \tag{9}$$

where $\cos\alpha'$, $\cos\beta'$, and $\cos\gamma'$ are the direction cosines of the Lorentz force.

Thus, in order to determine the critical value of the pinning force per unit volume at a given point in a superconductor at a given magnetic induction, one must divide the length of the vector connecting the origin with the energy surface (Eq. (8)) and aligned with the magnetic induction by the length of the vector to the width surface (Eq. (9)) and aligned with the Lorentz force, i.e., $F_P=U_0/L$. This procedure defines the operator Ξ introduced above. If the vector of the current density and magnetic induction vector are aligned with the principal directions in a material, the Lorentz force is also aligned with the principal direction orthogonal to the other two:

$$F_P^1=\frac{U^z}{L^y},\quad F_P^2=\frac{U^x}{L^z},\quad F_P^3=\frac{U^y}{L^x},\quad F_P^4=\frac{U^y}{L^z},$$

$$F_P^5=\frac{U^z}{L^x},\quad F_P^6=\frac{U^x}{L^y}. \tag{10}$$

The pinning surface is thus defined by six principal vectors and should be constructed in a space of double the original dimensionality. Each direction of the magnetic induction corresponds to a perpendicular cross section of the pinning surface. The axes of this cross section define pinning forces for all directions of the Lorentz force possible for this direction of the magnetic induction. A result of the calculations of several pinning surface cross sections are given in Fig. 2c. One can see that even in the simplest case being discussed, the shapes of the cross sections are nontrivial.

Figure 3a shows an example of the pinning surface cross section corresponding to an induction vector aligned with the y axis (this is the case, in particular, when an external field is perpendicular to the foil surface). The xz cross section is defined by

$$(F_P(\beta))^2=(F_P^4)^2\cos^2\beta+(F_P^3)^2\sin^2\beta. \tag{11}$$

Each radius of this cross section corresponds to a certain pinning force, but in a real experiment not all these forces can be realized because of the effects of so-called guided flux motion.

3.2. Guided flux motion in a superconductor with anisotropic pinning

Guided flux motion in a mixed state of type II superconductors was extensively studied as early as the late 1960s.[22,30] The transverse voltage in thin cold-rolled foils was measured in a perpendicular external magnetic field. The results were interpreted in terms of an excess pinning force in a certain fixed direction (perpendicular to the roll direction) and corresponding motion of fluxoids along the ''easy'' axis. But this model did not describe the mechanism of excessive pinning. Recently[31] guided flux motion has been detected in $YBa_2Cu_3O_{7-\delta}$, where a system of aligned twin boundaries played the role of barriers with excess pinning.

Figure 3 provides an interpretation of the guided motion, which naturally derives from the general concept of anisotropic pinning. Suppose that the transport current direction is such that the effective Lorentz force (1) aligned with the ab radius makes an angle β with the z axis. In the case of isotropic pinning, the onset of the flux motion occurs when the Lorentz force is larger than the pinning force. In the case of anisotropic pinning, when pinning forces are different for different directions, there are axes along which this force is so small that the Lorentz force projection (ad) exceeds the pinning force at a lower current than that at which the Lorentz force vector extends to the pinning surface ($ab<ac$).

As a result, the flow of fluxoids is directed along the critical projection of the Lorentz force, which leads to a deviation of the electric field vector from the current direction by angle Γ. The dashed line in Fig. 3a shows a calculated locus corresponding to the initiation of such a critical Lorentz force projection in the model of an ellipsoidal potential well (9). One can easily check that this curve is an ellipse with principal axes F_P^3 and F_P^4. A fluxoid array should be driven along the critical projection.

Figure 3b shows calculations of the angle between the directions of the fluxoid array motion and Lorentz force ver-

sus angle. Note that, when the force is aligned with a principal axis (in this case, x or z axis), the direction of the Lorentz force critical projection coincides with the Lorentz force, and $\Gamma=0$. Therefore, only in the principal directions is the pinning force exactly equal to the measured Lorentz force. Thus, the real critical current corresponding to the critical projection is less than or equal to the critical current calculated without taking the flux motion direction into account; nevertheless, it is possible to reconstruct the shape of the potential well using experimental data if the ''critical current'' corresponding to the critical Lorentz force and its angle with the electric field, $\Gamma=\arctan(E_\perp/E_\parallel)$, are measured simultaneously in a set of superconducting samples with various β. It clearly follows from Fig. 3a that the pinning surface is described by

$$F_P(\beta+\Gamma)=F_L(\beta)\cos\Gamma, \tag{12}$$

where $F_L(\beta)$ is the critical Lorentz force whose direction is determined by angle β.

The effects related to guided flux motion complicate the equation system (1)–(4) since, instead of the Lorentz force, its projection in the direction of flux motion must be included, and the pinning force should also correspond to this direction. In order to determine this direction, one should plot, using Eq. (3), the entire pinning surface cross section by the plane perpendicular to the magnetic induction vector.

3.3. Absolute values of the depth and width of the potential well

In principle, each of the six pinning forces can be derived from experimental data. One can easily check, however, that the following relation should hold:

$$F_P^1F_P^2F_P^3=F_P^4F_P^5F_P^6. \tag{13}$$

This condition does not allow one to determine the absolute values of the axes of the energy and dimension ellipsoids. But, since the forces are expressed in terms of energy and length, the principal axes of the energy and dimension surfaces can be normalized to one of these axes, for example, $L^x=a$. We have

$$U^y_{\max}=F_P^3a,\quad U^z_{\max}=F_P^5a,\quad U^x_{\max}=\frac{F_P^2F_P^3}{F_P^4}a=\frac{F_P^5F_P^6}{F_P^1}a,$$
$$L^y=\frac{F_P^5}{F_P^1}a,\quad L^z=\frac{F_P^3}{F_P^4}a. \tag{14}$$

The parameter a, which specifies the absolute scale of the system, cannot be derived from critical current measurements alone. In order to obtain absolute characteristics of the collective potential well, additional independent measurements are needed.

Equation (13) could be checked experimentally if there were a bulk anisotropic superconductor with the simple anisotropy described above. By cutting from this material three samples oriented in the principal directions and measuring critical currents in magnetic fields also aligned with the principal directions, one would obtain the six desired values of the pinning force not affected by the guided flux motion. Unfortunately, this sort of field experiment is not feasible in studying niobium-titanium foils.

TABLE I. Orientation of samples studied in experiments: ξ, θ, and φ (degrees).

$\xi=0°$		$\xi=15°$		$\xi=30°$		$\xi=60°$		$\xi=90°$	
θ	φ	θ	φ	θ	φ	θ	φ	θ	φ
0	90	0	90	0	90	0	90	0	90
90	–	90	–	90	–	90	–	90	–
45	0	45	0	45	0	45	0	45	0
0	45	0	45	–	–	0	45	–	–
45	90	45	90	–	–	45	90	45	90
–	–	–	–	30	35	30	35	–	–

4. EXPERIMENTAL RESULTS

4.1. Measurements of critical currents in external field

Samples were cut from cold-rolled foil of the NT-50 alloy (Nb plus 50 mass % of Ti) 10 μm thick produced in 1978 at the Ulba metallurgic factory. The foil was coated on both sides with a 1-μm layer of copper welded to the alloy in the process of pressing and rolling, which allowed us to fabricate good electrical contacts. We cut a set of ribbons (6×34 mm) oriented at angles $\xi=0°$, 15°, 30°, 60°, and 90° with respect to the roll direction from a 1-m long section of a piece of foil 200 m long and 80 mm wide. Thus we defined the transport current direction with respect to the preferred direction in the material. All ribbons were packed in one stack and machined in a conductor, namely, samples with necks 1.3 ± 0.05-mm wide were milled, and this helped in comparing experimental results and made accurate measurements of all sample cross sections unnecessary. By examining transverse and longitudinal polished cross sections of samples under a microscope, we checked that these cross sections were uniform along their lengths, and measurements of several samples of the same orientation taken from both the middle and ends of the foil section yielded the total experimental uncertainty (due to both inhomogeneity of the superconductor and differences in the sample widths), which was at most 8% throughout the magnetic field range studied.

The samples were soldered to stainless steel substrates and rigidly fixed to a probe by oriented holders. They were placed in the 41 mm diameter bore of a superconducting solenoid capable of generating a magnetic field of up to 12.5 T. The sample orientation in the external field was described by the angle θ between the field vector and foil plane, and the angle φ between the current and field projection in the sample plane. The sample orientations studied in the experiments are listed in Table I.

The rather large number of tested configurations was dictated by the need for sufficient statistics to check the model with five fitting parameters.

The experiment was conducted by measuring at the liquid-helium temperature (4.2 K) the current–voltage characteristics of samples with various orientations (various ξ) in a fixed set of magnetic fields ranging between 10 mT and

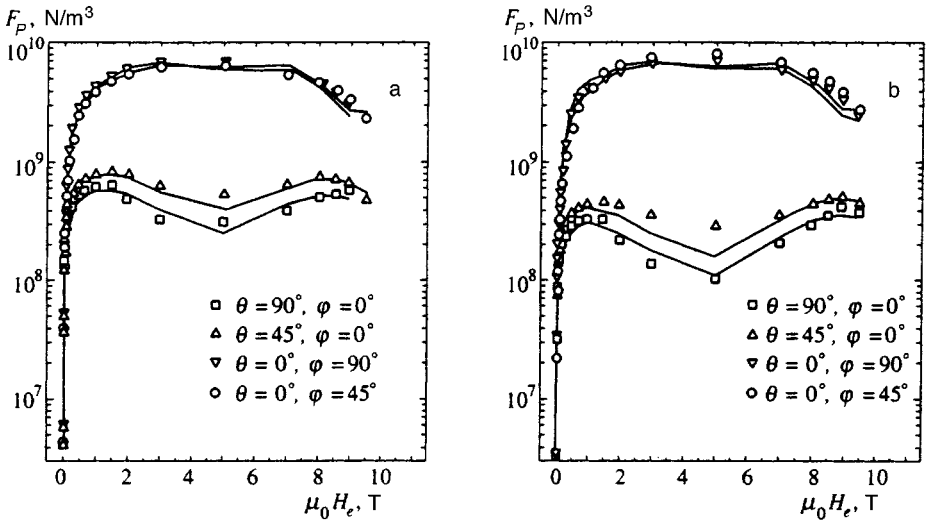

FIG. 4. Comparison between experimental measurements (points) of the pinning force in anisotropic niobium–titanium foil and calculations (solid lines) based on the parameters of the collective potential well given in Fig. 6 for samples cut at an angle of (a) 15° and (b) 60° with respect to the roll direction.

H_{c2}, which was determined to be 10.8 T for the foil (at a current of 1 mA and electric field of 1 μV/cm). In this paper, we discuss only a fraction of the information obtained, namely the critical current in a longitudinal electric field of 1 μV/cm as a function of the external magnetic field orientation and current direction in the foil.

The goals of our experiments were the following:

(a) To obtain reliable experimental information on the critical current as a function of external field direction and current orientation with a view to determining parameters of the anisotropic pinning model and verifying that the current-carrying capacity of a material really depends on both the Lorentz force and magnetic induction directions in the material.

(b) To record the shapes of the maxima in the current-carrying capacity at low fields and certain external field orientations for comparison with calculations. These maxima were detected many times by other researchers in various materials, including the niobium–titanium foil,[9] when the external field was aligned with the foil plane, but have not been explained.

Experimental data are given in Fig. 4 with fits to the model described above. In order to limit the number of plots in the paper, we have included only a small fraction of the experimental data, namely the results which demonstrate the best and worst agreement with the model calculations.

Let us list the main experimental results that are by and large in agreement with the anisotropic pinning model.

(1) The tested material (thin cold-rolled NT-50 foil) is a material with strongly anisotropic pinning. The difference between the pinning force densities at certain orientation reaches two orders of magnitude for a given external field strength. The highest current-carrying capacity was detected when the magnetic field was aligned with the foil plane, and the lowest, when the field was normal to the foil.

(2) The pinning force per unit volume strongly depends on the magnetic induction inclination to the foil material (Fig. 4), even when the Lorentz force direction is the same, which is clearly seen in Figs. 4a and 4b.

(3) The pinning force density depends heavily on the Lorentz force direction in the foil material, even when the strength and direction of the external field are the same. In order to check this, one should compare measurements for $\xi = 15°$ and 60° given in Fig. 4.

(4) At certain orientations of the external field (along the foil plane), with a field strength comparable to the local induction in the material due to characteristic transport currents, i.e., below 0.2 T, maxima in the critical current versus magnetic field were detected (Fig. 5).

4.2. Measurements of transverse electric field

As was noted above, in type II superconductors with anisotropic pinning, a flow of fluxoids can deviate from the direction perpendicular to the current, thereby inducing a transverse component of electric field. By measuring the angle between the current and electric field in an external magnetic field notably higher than the field generated by the current, one can plot a 2D cross section of the pinning surface perpendicular to the field direction (under the assumption that the critical projections are in the plane of this cross section). Like other experiments in this series, the experiment was carried out on NT-50 alloy cold-rolled foil. We cut ribbons from the foil area about its middle line with dimensions of 5×30 mm oriented at angles $\xi = 0°$, 10°, 15°, 20°, 30°, 40°, 60°, and 90° to the roll direction.

The samples were soldered to stainless steel plates and affixed to a probe such that the magnetic field generated by the superconducting solenoid was perpendicular to their plane. With this alignment, detection of the transverse electric field is easier. In order to measure the longitudinal and transverse voltages simultaneously, we fabricated a device with four spring-loeaded potential contacts located at the corners of a square with a 3.5-mm diagonal, one of the diagonals being aligned with the sample axis so that is coincided with the current direction. The lack of a spurious signal due to positioning errors in the contact positions was tested at zero magnetic field in samples in the normal state. In this case, the transverse signal was within 2% of the longitudinal signal, and this result was satisfactory since the detected

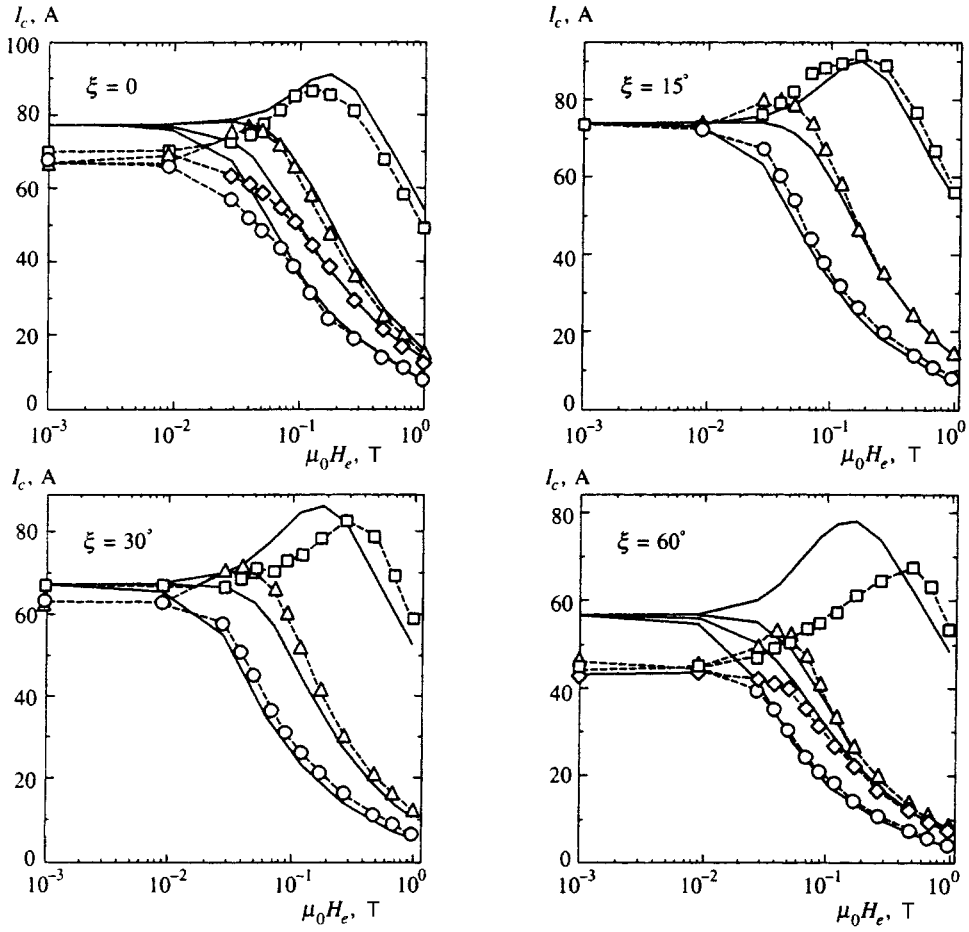

FIG. 5. Comparison between experimental data and calculations of critical current versus magnetic field in the range of low magnetic fields. The dashed lines and symbols show experimental data, and solid lines show calculations using the parameters of the collective potential well given in Fig. 6: circles) $\theta=90°$, $\varphi=0°$; rhombs) $\theta=45°$, $\varphi=90°$; squares) $\theta=0°$, $\varphi=90°$; triangles) $\theta=45°$, $\varphi=0°$.

transverse voltage in a magnetic field was usually much higher. Given the large magnitude of the studied effect, we could ignore the Hall voltage, which was within the experimental uncertainty. This device allowed us to save time by avoiding the complicated procedure of potential contact alignment before measuring a new sample.

In magnetic fields ranging between 0.5 and 10 T, in which the field generated by the current could be neglected and the induction in the superconductor was aligned with the external magnetic field, we recorded direct and reverse current–voltage characteristics in the voltage range 0.01 to 10 μV using our computerized apparatus. Concurrently we plotted the electric field transverse component versus the longitudinal component on an X-Y recorder. When the current was increased and decreased, the recorder plotted identical straight lines. This was an indication of the direct proportionality between the transverse and longitudinal components, hence the constant trajectory of the magnetic flux motion. The slope of these lines corresponded to the angle between the electric field and current. It strongly depended on the angle between the roll direction and current. In samples whose axis was aligned with the roll direction or was perpendicular to it, no transverse component of the electric field was detected, which supported our assumption that there were at least two principal directions of the pinning surface in the material, and that they coincided with the principal directions in the foil.

4.3. Data processing

In the case of arbitrary sample tilt with respect to an external magnetic field, the pinning force is controlled by several parameters that define the pinning body (Table II), so statistical methods are needed in the data processing. The situation is more complicated in weak external fields, where the field generated by the current affects the magnitude and direction of the magnetic induction in the sample, and a self-consistent problem of calculating the current density must be solved.

The aim of the data processing was to determine five pinning parameters, namely $U^x(B)$, $U^y(B)$, $U^z(B)$, $L^x(B)$, and $L^z(B)$, which were treated as independent values in our

TABLE II. Orientation of samples and relevant model parameters affecting the pinning force.

Orientation of current and external field, degrees		Relevant model parameters affecting the pinning force density in high magnetic fields	
θ	φ		
0	90	U^x, U^z	L^y
90	–	U^y	L^x, L^z
45	0	U^x, U^y, U^z	L^x, L^z
0	45	U^x, U^z	L^y
45	90	U^x, U^y, U^z	L^x, L^y, L^z
30	35	U^x, U^y, U^z	L^x, L^y, L^z

model, to derive from them self-consistent distributions of the current density and magnetic induction for all orientations and fields used in the experiments, and to compare calculations to measurements of the currents. We hoped that the description of the angular dependencies of the potential well depth and width with ellipsiods would be in reasonable agreement with reality.

4.3.1. Effect of current density nonuniformity in weak external field. In order to determine the critical current density $j_c(\mu_0 H_e)$ from the measured critical current $I_c(H_e)$, the critical current is usually divided by the sample cross section. This procedure is not always justified, since according to the critical state model, the current density in the sample is correlated with the local induction, and the latter is usually nonuniform because of the field generated by the transport current. For samples of isotropic materials calculating the current density by dividing the total current by the cross section area is correct only in the case of a sufficiently high external field. Then an increase in the current density in regions where the induction is lower is exactly compensated by its decrease in the region where the induction is higher. This was analytically demonstrated for the case of an isotropic superconductor in the shape of a cylinder, with the current density being a power function of the induction.[18]

We assumed and then checked in a numerical calculation using the algorithm described below that, in the studied case of an anisotropic superconductor with a rectangular cross section and a realistic function $j_c(B)$, one can apply this technique to fields higher than 0.1–1.0 T (depending on the field tilt with respect to the foil), when the angle between the external field and current is sufficiently large (more than 10°). If the sum of the external field and the field due to the current is sufficiently small, a peak in the current density is formed at this point, and in the major part of the cross section the total induction is greater than the external magnetic field induction. Therefore, the critical current in the sample is considerably lower than the current density corresponding to the external field times the cross sectional area.

The parameters of the anisotropic pinning model throughout the field range between zero and H_{c2} (i.e., the parameters U^x, U^y, U^z, L^x, and L^z of the anisotropic potential well versus external field) should be obtained by statistically processing all experimental data measured at all orientations of the magnetic field for all samples. In order to estimate the pinning parameters and solve the self-consistent problem in small fields, we had to use values of the critical current density in weak fields extrapolated to this range by the procedure described below. If we assume that the pinning force density as a function of magnetic field in the weak-field regime can be expressed as

$$F_P \propto b^q (1-b)^s, \tag{15}$$

where q and s are parameters that depend on the properties of pinning centers, and $b = B/B_{c2}$ is the reduced magnetic field, it is natural to expect that in the anisotropic case the parameters in Eq. (15) depend on the field direction and Lorentz force orientation.

Analysis of the measured critical current versus field has confirmed our assumptions. Irrespective of the external field orientation, for all samples over a certain range of weak reduced fields ($b \sim 0.01-0.2$), one can select linear sections on curves of the critical current versus field plotted on a log–log scale on which Eq. (15) holds to fair accuracy (i.e., $I_c \propto b^{q-1}$). In weaker external fields, the field due to the transport current begin to make itself felt, and at higher magnetic fields, it is the peak effect, which leads in both cases to deviations of experimental curves from a power law. We assumed that the resulting curves could be extrapolated to very weak external fields, and that the required curve of $j_c(B)$ in this range for calculations could be obtained. The curve was extrapolated using the least squares method by minimizing the discrepancy factor across all experimental curves, and the optimal values of extrapolation parameters were derived with the key parameter q being varied between 0.378 and 0.709. In the calculations of the pinning parameters given below, we described the function $j_c(\mu_0 H_e)$ in an external field $\mu_0 H_e \geqslant 1$ T by $I_c(H_e)/S$, and for $\mu_0 H_e < 1$ T by the extrapolated critical current through the cross sectional area.

It is significant that all approximations used in our calculations yield a critical current density that decreases monotonically with magnetic induction. The strong dependence of the critical current on the orientation allows us to ascribe maxima in the critical current at low magnetic fields for some orientations to features of the magnetic induction distribution over the sample cross section due to the field generated by the transport current, which is also highly nonuniform across the sample cross section.

4.3.2. Calculation of model parameters. We constructed twenty curves of critical current versus field, $I_c = I_c(H_e, \xi, \theta, \varphi)$, corresponding to the angles ξ, θ, and φ listed in Table II. An independent calaculation was performed for each value of external field and all mutual orientations of current and field. The model parameters were derived by minimizing the discrepancy factor R:

$$R = \frac{1}{N-\nu} \sum_k^N \left(\frac{I_{c,k}^{\text{cal}} - I_{c,k}^{\text{exp}}}{I_{c,k}^{\text{exp}}} \right)^2, \tag{16}$$

where N is the number of experiments, $I_{c,k}^{\text{exp}}$ is the critical current measured for the kth orientation, $I_{c,k}^{\text{cal}}$ is the critical current calculated by one of the models for the kth orientation, and ν is the number of adjustable model parameters. The regression procedure was performed assuming an equal contribution from each experimental point (measured critical current) to the discrepancy factor.

The calculated parameters are plotted in Fig. 6. As mentioned above (see Eqs. (13) and (14)), given six main values of the pinning force, one can determine only five relative values of the potential well parameters, and one of its dimensions is selected as a normalization factor. The shapes of curves of these parameters versus magnetic induction are changed by the normalization to different principal radii of the dimensional ellipsoid, because these principal radii can variously depend on the magnetic induction.

Nevertheless, the data plotted in Fig. 6 are sufficient to derive complete information about pinning in the material. In order to determine the principal values of the pinning force expressed in N/m^3, one must multiply the ratio between re-

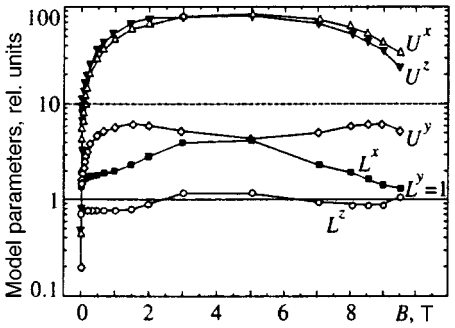

FIG. 6. Parameters of the collective potential well calculated using measurements of the anisotropy of the current-carrying capacity of niobium–titanium foil.

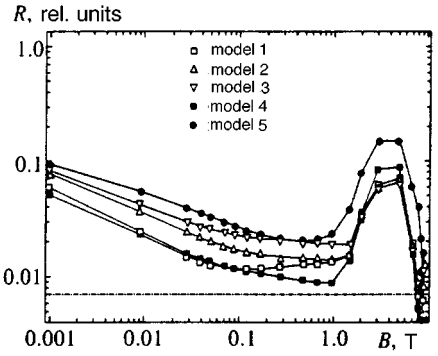

FIG. 7. Qualitative estimate of reliability of different models based on the R-criterion. The dot-and-dash line shows the average experimental uncertainty.

spective energies and dimensions given in Fig. 6 by 10^8. The normalization for the data in Fig. 6 ($L^y=1$) corresponds to the case of the flattest curves of collective potential well parameters plotted as functions of the field, with minimal nonmonotonic sections. Note that the density of grain boundaries should be the highest in the direction of L^y, so that, in our opinion, the potential well width should remain constant up to a very high magnetic induction at which the average separation between fluxoids can become smaller than the distance between pinning centers.

Figure 6 indicates that the collective well depth is essentially constant for all induction directions in the foil plane. For the induction direction normal to the foil plane, the well is an order of magnitude shallower. Typical well widths for a Lorentz force in the roll direction are several times larger than in perpendicular directions almost throughout the induction range studied, but starting with 5 T, this difference becomes smaller and vanishes around the critical field. In our opinion, this behavior is in fairly good agreement with the assumption that the pinning centers in the cold-rolled foil are boundaries of grains strongly flattened and elongated in the roll direction. If the difference between the x and z axes of the energy and dimensional ellipsoids is presumed to be physically insignificant, the number of fitting parameters can be reduced to three.

4.3.3. Comparison of models. In order to check how good the description of the current-carrying capacity is as a function of orientation in terms of the collective potential well parameters, we compared the statistical parameters of our model to those derived from alternative simplified models. In each case, we used the same set of experimental data and performed the procedure of nonlinear regression. Equation (16) allowed us to perform a correct comparison among models with different numbers of fitting parameters using the R-criterion, whose sense is similar to that of the χ^2-criterion. The following models were compared.

1) The model of anisotropic pinning described above with the field-dependent depth and width parameters of the collective potential pinning well. The variable parameters were the depth and width of the well, U^x, U^y, U^z, L^x, and L^z ($\nu=5$).

2) A similar model, but by taking into account the symmetry of the pinning center distribution in the foil, the number of fitting parameters is reduced to four, since $U^x=U^z$ ($\nu=4$).

3) A model similar to the previous one, but with an additional symmetry element and the number of fitting parameters reduced owing to the additional relation $L^z=L^y=1$ ($\nu=3$).

4) A model similar to the first, but taking into account the guided flux motion in an anisotropic superconductor. When the guided flux motion is taken into account, the measured critical current is lower in almost all cases than the value obtained in the expected Lorentz force direction. In the case of the simplest shape of the potential well, namely elliptical, the guided flux motion can be taken into account in analytic form, which leads only to minor modifications in theoretical formulas, and the number of variable parameters remains the same ($\nu=5$).

5) An alternative model in which the pinning force depends only on the direction of the Lorentz force in the material. This model is equivalent to the well-known extension of the Bean model.[14,15] We took as model parameters the three induction-dependent principal pinning forces along the three principal directions in the material, and assumed that for all other orientations of the Lorentz force the pinning force is determined by the vector of the ellipsoid defined by the three principal axes ($\nu=3$).

The results of the five models are compared in Fig. 7 in the form of the discrepancy factor R versus field for various models. The experimental uncertainty (due to the spread of the position-dependent foil parameters, errors in sample mounting, the small spread of the sample widths) derived from measurements of the critical current of a set of varied samples for the same orientation of the external field is shown in this graph by the dot-and-dash line. The smaller R for the selected model and the closer its value to the experimental uncertainty, the better the description of experimental data by this model. It is clearly seen that the ''generalized Bean model'' has a discrepancy factor considerably (almost three-fold) higher than this factor for other versions of the anisotropic pinning model, which takes account of the magnetic induction direction in the material. Even the model describing the case of very symmetrical anisotropic pinning (the third one) with the same number of variable parameters

($\nu=3$) yields a more accurate description of experimental data.

Nevertheless, the statistical error associated with all models is considerably larger than the experimental uncertainty, and in the range of 2 to 7 T the discrepancy factor is enhanced for all models. (The rise in the discrepancy factor when the field drops below 0.1 T can be ascribed to the lack of accurate measurements of the critical current densities, and instead of these data we took extrapolations from the range of higher fields, which, naturally, led to larger errors.) It would be natural to express, in response to such discrepancies, doubt that the suggested ellipsoidal shapes of the energy and dimension surfaces are justified, but we think that the real cause is, most probably, the inadequacy of the critical current evaluation criterion related to an arbitrary level of electric field. Indeed, this choice of critical current definition is based on tradition alone. An alternative criterion based on a constant effective resistance would change the curves significantly, especially at intermediate fields, where the difference between critical currents in differing orientations is greatest, but this criterion also cannot be rigorously justified. We hope to eliminate this uncertainty after the planned processing of current–voltage characteristics.

5. COMPARISON BETWEEN CALCULATIONS AND MEASUREMENTS; DISCUSSION

In order to compare experimental data to calculations and take into account the field generated by the transport current, the problem of self-consistent current and induction distributions in the sample cross section was solved numerically throughout the field range studied. We applied the method of successive approximations to the rectangular cross section divided into 500–1000 elements. The pinning force, with due account of the local induction vector and Lorentz force direction, was applied determined for each element. Usually a solution was found after twenty to thirty iterations. The stability of the result against changes in grid resolution was checked.

The calculations are compared to experimental data in Figs. 4 and 5. Figure 4 compares calculations and measurements of the pinning force for various orientations of the magnetic field in samples cut at angles of 15° and 60° with respect to the roll direction. For most orientations, the agreement between model calculations and experimental measurements is satisfactory. Figure 5 compares calculations and measurements of critical currents in weak external fields, since correct measurements of the pinning force in these fields are difficult, and features due to the nonuniformity of the current density distribution show up precisely in the behaviore of $I_c(H_e)$. The calculations faithfully reproduce all features of these curves in weak fields, such as maxima in the current for some field orientations and the absence of maxima for others.

An unexpected effect for us was a maximum in the curve corresponding to the field orientation at 45° with respect to the current in the plane perpendicular to the foil ($\varphi=0°$ and $\theta=45°$). We attribute this maximum to features in the current distribution at small angles between the magnetic induction and current, rather than material anisotropy. Most probably, this maximum is a rudiment of the giant maximum expected in the case when the external field is aligned with the current even in an isotropic configuration, but in our calculations the remnant of this maximum persists at an angle θ ranging only to 40°. We tentatively ascribe some quantitative differences between calculations and measurements at fields of less than 0.1 T (Fig. 7) to either poor accuracy of $j_c(B)$ extrapolation to low fields, which leads to a large uncertainty in the current for a small error in the approximation parameters, or a fundamental flaw in this procedure, because Eq. (15) fails at very low fields.

Figure 8 shows calculations of the self-consistent critical current density distribution in the sample cross section, which elucidates the nature of maxima in the field parallel to the foil plane. The relatively uniform current distribution in a moderately high external field (0.5 T) becomes more and more nonuniform with decreasing field, since the field generated by the current reduces the induction on one edge of the cross section and gives rise to a peak in the current density. This peak generates a fairly large field component perpendicular to the sample plane in the foil cross section with a large ratio between its sides, and this process dominates, since in a transverse field the pinning force is an order of magnitude lower that in a parallel field with the same induction. This then produces the maximum in $I_c(H_e)$.

Figure 6 shows how the potential well depth depends on field orientation. Available data on the characteristic dimension L^y as a function of magnetic field are insufficient for determination of the specific binding energy between magnetic flux and pinning centers. If we assume that L^y is constant with the magnetic induction, i.e., the density of pinning centers is higher than the fluxoid density throughout the field range studied, and, on the other hand, take into consideration that the separation between pinning centers should be larger than the coherence length, we can obtain a very rough estimate $L^y\approx\sqrt{\phi_0/B}$ at $B\approx10$ T. Thus, $L^y\approx7$ nm, and the binding energy at an induction of 5 T is $U^x\approx U^z\approx50$ J/m^3, $U^y\approx2.5$ J/m^3.

The comparison in Fig. 9c between the ratio of the pinning force perpendicular to the roll direction and that parallel to this direction, and the ratio between widths of the collective potential well (Fig. 6) derived from the entire set of critical current measurements, leads us to conclude that the anisotropy taking place in the case of this orientation is uniquely related to the difference between the slopes of potential well walls. The peak effect detected around B_{c2} is due to the increase in the wall slope, rather than the increase in the binding energy between magnetic flux and pinning centers.

Measurements of transverse electric field in magnetic fields perpendicular to the foil plane allow us to reconstruct the real shape of the cross section of the pinning body. Figure 9 shows reconstructed shapes of pinning surface cross sections. Throughout the field range studied, they are very similar to the cross section calculated in Sec. 3.2. The suggestion of an ellipsoidal approximation for the depth and width of the collective potential well in a niobium–titanium foil for this magnetic field orientation is probably quite realistic. Naturally, the shape of this potential well cannot be universal. For example, in HTSC the presence of several

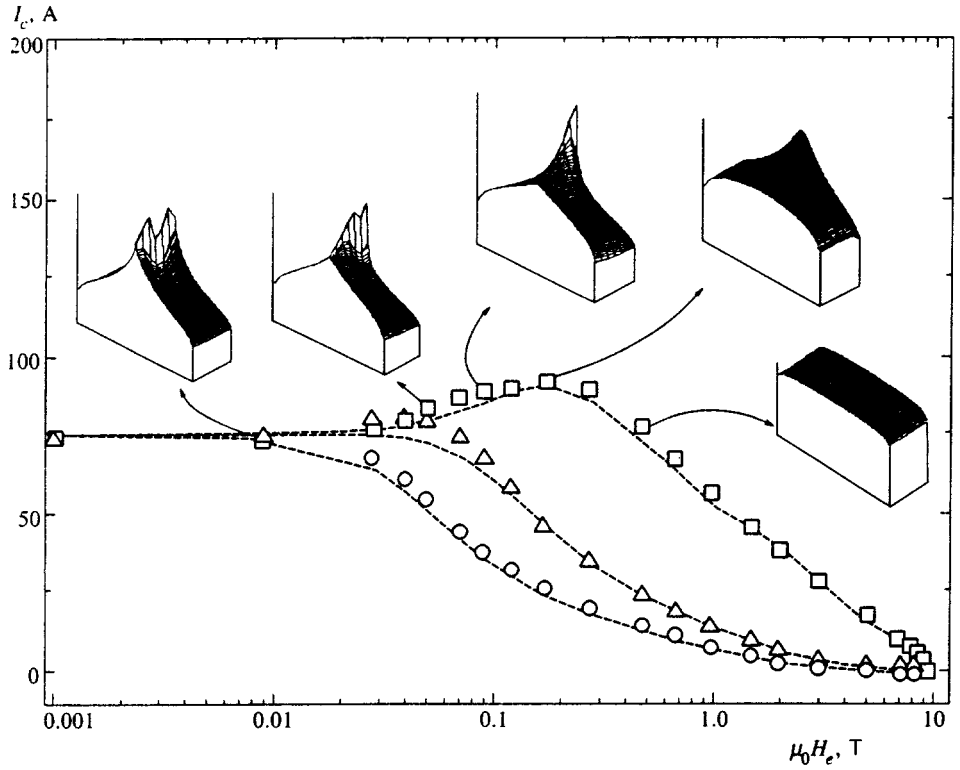

FIG. 8. Diagram explaining initiation of critical current maxima in small magnetic fields parallel to the roll direction in niobium–titanium foils with $\xi = 15°$; ○—$\theta = 90°$, $\varphi = 0°$; △— $\theta = 45°$, $\varphi = 0°$, □—$\theta = 0°$, $\varphi = 90°$. The inserts show distributions of the current density in the sample cross section corresponding to different points on the curves of critical current versus field (for a better view, the dimension of distribution diagrams along the larger axis is compressed considerably because the ratio between the cross section sides is 130:1).

ensembles of pinning centers with different orientations, which leads to several sharp peaks in the current-carrying capacity, should correspond to a very peculiar pinning surface shape resembling a star. We hope that a study of such cross sections in various materials can be helpful in elucidating details of their structures and distributions of real pinning centers, and in particular, that the observed features of measured cross sections enable us to improve upon the model of magnetic flux pinning in niobium–titanium alloys.

6. CONCLUSIONS

The proposed geometrical interpretation of the macroscopic pinning force as a surface pierced by the Lorentz force, which means a transition from the superconducting to resistive state, has allowed us to construct a model of anisotropic pinning whose difference from the generalized Bean model is that it takes into account the dependence of the pinning force not only on its direction, but also on the magnetic induction direction. A simple algorithm describes an operator that transforms the induction vector and Lorentz force in the counteracting pinning force for the case in which the width and depth of the collective pinning potential well can be described by triaxial ellipsoids. The suggested model yields a natural interpretation of guided fluxoid motion and predicts the angle between the current and electric field in the resistive state of an anisotropic superconductor.

An algorithm has been developed for calculating the distribution of the self-consistent current and magnetic induction in the superconductor cross section for arbitrary magnetic field and current orientations, with due account of the real anisotropy of each cross section element.

The current-carrying capacity of a cold-rolled niobium–titanium foil has been studied for various magnetic field orientations as a function of the current direction in the coordinate system attached to the material, and transverse electric fields in the foil resistive state in magnetic fields perpendicular to the foil plane have been measured. A technique for reconstructing the real shapes of pinning body cross sections from measurements of the angle between current and electric field in the resistive state has been suggested, and the shapes of the pinning body cross sections for one induction orientation have been constructed.

The comparison between experimental measurements and model calculations demonstrates that the approximation of the pinning potential well by two ellipsoids is satisfactory for the niobium–titanium foil. In HTSC materials with several systems of sharply oriented ensembles of pinning centers, one should expect a pinning surface of very complex shape, because both the depth and width of the potential well should have star-like shapes formed by several highly elongated intersecting ellipsoids.

Relative depths and widths of the collective potential well when both the magnetic induction and Lorentz force directions coincide with the principal directions of the niobium–titanium foil have been derived from experimental data. The peak-effect in fields close to the critical value detected for some orientations must be ascribed to changes in the steepness of the potential well walls, rather than the potential well depth.

By solving the problem of self-consistent distributions for the magnetic induction and current density in samples with due account of the field generated by the transport current, we could reproduce in our calculations of the critical

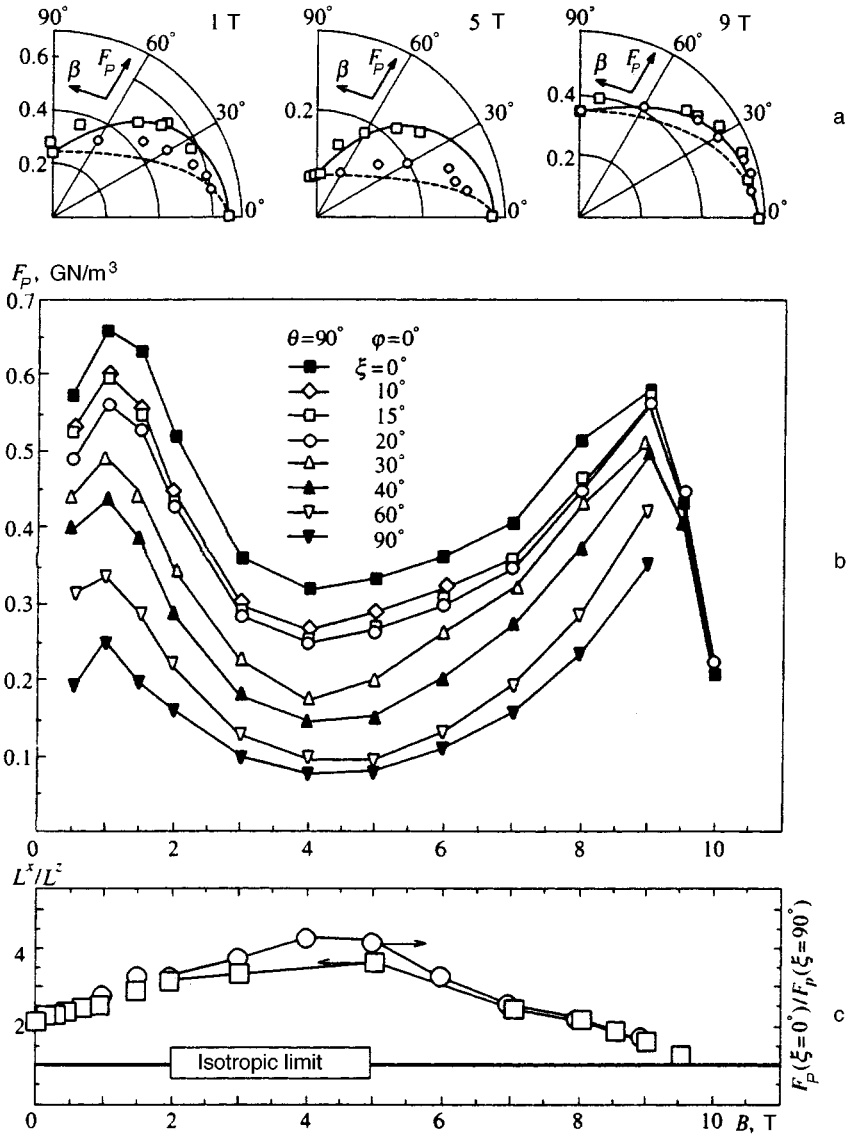

FIG. 9. Reconstruction of the pinning body shape. a) Typical reconstruction for three values of external field demonstrating satisfactory agreement between experimental data and calculations. The dashed lines and circles show the calculations and measurements of maximum current for samples cut at various angles with respect to the roll direction. The solid lines show the expected pinning body cross section, and squares show reconstructions of these sections from experimental data. b) Pinning force versus field for samples cut at various angles with respect to roll direction in a magnetic field perpendicular to the foil plane. c) Comparison between the ratio of pinning forces along the two principal axes in the foil plane and the ratio of potential well dimensions for the same directions.

current versus external field some subtle features, including maxima in weak external fields parallel to the sample plane.

The model can be used in qualitative description of pinning in practical superconducting materials, as well as in quantitative description of processes in superconducting wires used in magnetic fields at varying orientation.

The authors are grateful to A. P. Zhernov, K. A. Kikoin, T. D. Aksenova, and E. P. Krasnoperov for helpful discussions. This work was supported by the Russian Fund for Fundamental Research (Project 96-02-18693), the Council on Superconductivity (Project 93176), and Board of the Kurchatov Institute (project for support of young scientists No. 96-40).

[1] C. J. Bergeron, Jr., Appl. Phys. Lett. **3**, 63 (1963).
[2] J. R. Clem and S. Yeh, J. Low Temp. Phys. **39**, 173 (1980).
[3] G. Blatter, M. V. Feigel'man, V. B. Geshkenbein, A. I. Larkin, and V. M. Vinokur, Rev. Mod. Phys. **66**, 1125 (1994).
[4] E. H. Brandt, Rep. Prog. Phys. **58**, 1465 (1995).
[5] V. M. Boroditch, A. C. Combarov, V. Ya. Filkin *et al.*, in *Low Temperature Physics-LT8* (Proc. of Int. Conf. LT8, London, 1962), R. O. Davies (ed.), Butterworth, London (1963), p. 348.
[6] R. R. Hake, D. H. Leslie, and C. G. Rhodes, *ibid.*, p. 342.
[7] K.-P. Jungst, IEEE Trans. Magn. **MAG-11**, 340 (1975).
[8] E. W. Collings, in *Applied Superconductivity, Metallurgy and Physics of Titanium Alloys*, Vol. 2, Plenum Press, London–New York (1986), p. 100.
[9] B. ten Haken, L. J. M. van der Klundert, V. S. Vysotsky, and V. R. Karasik, IEEE Trans. Magn. **28**, 755 (1992).
[10] H. Kupfer and T. Matsushita, J. Appl. Phys. **63**, 5060 (1988).
[11] S. J. Williamson and J. K. Furdina, Phys. Lett. **21**, 376 (1966).
[12] C. M. Friend and D. P. Hampshire, in *Applied Superconductivity* (Proc. EUCAS-93), H. C. Freyhardt (ed.), DGM, Inform. Gesellschaft L. B. H. Verlag (1993), Vol. 1, p. 701.
[13] A. M. Campbell and J. E. Evetts, *Critical Currents in Superconductors*, Taylor and Francis Ltd., London (1972).
[14] E. M. Georgy, R. B. van Dover, K. A. Jackson, L. F. Shneemeyer, and J. V. Waszczak, Appl. Phys. Lett. **55**, 283 (1989).
[15] F. M. Sauerzopf, H. P. Wiesenger, and H. W. Weber, Cryogenics **30**, 650 (1990).
[16] H. P. Wiesinger, F. M. Sauerzopf, and H. W. Weber, Physica C **203**, 121 (1992).
[17] G. Blatter, V. B. Geshkenbein, and A. I. Larkin, Phys. Rev. Lett. **68**, 875 (1992).

[18] E. Yu. Klimenko and V. G. Kon, in *Superconductivity* (Proceedings of Conference on Technical Applications of Superconductivity, Alushta-75), Atomizdat, Moscow (1977), Vol. **4, p.** 114.
[19] V. R. Karasik and V. G. Vereshchagin, Zh. Éksp. Teor. Fiz. **59**, 36 (1970) [Sov. Phys. JETP **32**, 20 (1970)].
[20] C. P. Bean, Phys. Rev. Lett. **8**, 250 (1962).
[21] E. Yu. Klimenko and N. N. Martovetsky, IEEE Trans. Magn. Mater. **28**, 843 (1992).
[22] A. K. Niessen, J. van Suchtelen, F. A. Staas, and W. D. Druyvesteyn, Phys. Lett. **13**, 293 (1965).
[23] E. J. Kramer, J. Appl. Phys. **44**, 1360 (1973).
[24] J. Lowell, J. Phys. F: Metal Phys. **2**, 547 (1972).
[25] A. M. Campbell, J. Phys. C: Sol. St. Phys. **4**, 3186 (1971).
[26] A. I. Larkin, Zh. Éksp. Teor. Fiz. **58**, 1466 (1970) [Sov. Phys. JETP **31**, 784 (1970)].
[27] A. I. Larkin and Yu. N. Ovchinnikov, J. Low Temp. Phys. **34**, 409 (1979).
[28] C. B. Nunes and D. G. Pinatti, Supercond. Sci. Technol. **5**, S256 (1992).
[29] E. H. Brandt, Phys. Rev. Lett. **57(11)**, 1347 (1986).
[30] A. K. Niessen and C. H. Weijsenfeld, J. Appl. Phys. **40**, 384 (1969).
[31] V. N. Morgoon, V. A. Shklovskij, V. Bindilatti *et al.*, Czech. J. Phys. **46**, 1751 (1996).

Translation was provided by the Russian Editorial office.

Competition between face-centered cubic and icosahedral cluster structures

R. S. Berry

University of Chicago, Chicago, IL 60637, USA

B. M. Smyrnov, and A. Yu. Strizhev

High-Temperatures Institute, Russian Academy of Sciences, 127412 Moscow, Russia
(Submitted 20 April 1997)
Zh. Éksp. Teor. Fiz. **112**, 1082–1090 (September 1997)

The binding energy of atoms in icosahedral and face-centered-cubic clusters is calculated numerically for pairwise Morse-potential interactions between atoms and for clusters containing from 561 to 923 atoms, which corresponds to gradual filling of the sixth layer of the icosahedral cluster. Perturbation theory is used to calculate the cluster binding energy, in which the small parameter is the ratio of the interaction energy between non-nearest neighbor atoms to the interaction energy between nearest neighbors. Values of the Morse interaction potential parameter are found for which the energies of clusters with different structures coincide. Under the conditions used in these computations, the strain energy of a cluster can be neglected. Although the contribution of the interaction energy between non-nearest neighbors to the total cluster energy is small, it turns out to be important in finding the level crossing parameter. [S1063-7761(97)02009-X]

1. INTRODUCTION

Two-hundred thirty crystal structures are known to occur naturally at low temperatures.[1–3] The variety of cluster structures is still more extensive. For example, condensed-matter systems consisting of atoms with pairwise interactions form close-packed crystal structures, which include hexagonal and face-centered cubic (*fcc*) structures.[3,4] Clusters of atoms with pairwise interactions can condense into these structures as well, but they also can form icosahedra.[5] Geometrically, the icosahedron has a surface consisting of 20 equilateral triangles. The internal atoms of an icosahedral cluster have 12 nearest neighbors, like a close-packed structure in which all the distances between nearest neighbors are the same. However, in an icosahedral cluster the distance between nearest neighbors belonging to the same layer exceeds the distance between nearest neighbors belonging to adjacent layers by roughly 5%. Therefore, the structure of an icosahedron is unsuitable for macroscopic systems. However, for systems made up of a finite number of bound cluster atoms it can turn out to be preferable at low temperatures.

The advantage of the icosahedral structure compared to the close-packed structure lies in the larger number of bonds between nearest neighbors. Therefore, this structure will be energetically more favorable for small clusters. In particular, for clusters of inert gases a transition from an icosahedral structure to an *fcc* structure occurs when the numbers of atoms in the cluster reaches $n=1000-3000$.[6–8] For clusters of atoms interacting via a Lennard–Jones potential, the transition from the icosahedral structure to a cuboctahedron—one of the *fcc* structures—occurs when the number of atoms in the cluster satisfies $n\sim 10^4$ (Refs. 9,10), while for clusters with short-range interactions between atoms the transition between the icosahedral and *fcc* structures takes place for $n=200-500$.[11,12] This extreme sensitivity of the transition region to the type of atomic interaction requires a more careful analysis of the competition between these structures. In this paper we carry out this analysis using the method developed in Ref. 13 for a Morse-potential interaction between atoms. We then illustrate the competition between structures by using a ''phase diagram'' (see Ref. 13) on which we plot the Morse-potential parameter at which the *fcc* and icosahedral structure energies cross versus the number of atoms in a cluster. Constructing this phase diagram is in fact the main task of this article.

Note that we have excluded clusters with hexagonal structure from the discussion. This structure competes with the *fcc* structure for small-sized clusters, when the icosahedral structure is preferred, but in the range of atom numbers where the icosahedral and *fcc* structures compete, both these structures characteristically bind atoms to the cluster with higher binding energies than the hexagonal structures do.[14,15] In what follows we identify as ''icosahedral'' those clusters whose core has the structure of an icosahedron, although their surface occupied layers can have either *fcc* or icosahedral structures,[6,7,16] depending on the degree of occupation of the layer. We specify the structure of the surface layer of atoms by choosing the structure with the larger atomic binding energy for the given number of atoms.

A cluster with *fcc* structure grows by filling its planar facets.[14,17] Such structures come in pairs, depending on whether the core of the cluster does or does not contain a central atom. Of these two possible structures for a cluster we choose the one with the larger atomic binding energy for the given number of atoms in the cluster. Thus, when we compare *fcc* structures with icosahedra, we first choose the optimal configuration of atoms among the *fcc* structures, i.e., the one that ensures the maximum binding energy. In reality, each of the optimal atomic configurations will correspond to the maximum number of bonds between nearest neighbors.

We then compare the energies of optimal configurations of atoms with *fcc* and icosahedral structures for clusters with

numbers of atoms $n=561-923$, which corresponds to gradual filling of the sixth layer of the icosahedral structure. This comparison allows us to understand the distinctive features of the competition between these structures.

Note that the variety of crystal structures observed in nature is determined by the way the type of interactions between atoms differs from system to system. The assumption that this interaction is pairwise for a system of bound atoms in which the atoms retain their individuality is correct if the interaction potential between two atoms does not depend on the positions of the surrounding atoms of the system. This will be the case when the interaction potential between two atoms is small compared to characteristic single-atom quantities, in particular the ionization potential of an atom. Hence, our analysis will apply primarily to clusters consisting of inert-gas atoms or gas molecules. In other types of interactions the atoms lose their individuality when they enter the cluster, which complicates the analysis of the cluster structures. However, even the simple case under discussion here, i.e., pairwise atomic interactions, admits multiple cluster structures and their hybrids, such that the optimal cluster structures have *fcc* or icosahedral cores.

2. ENERGETICS OF LARGE CLUSTERS

The Morse-potential interaction between two atoms has the following form as a function of the interatomic distance R:

$$U(R)=D\{\exp\,[2\alpha(R_e-R)]-2\exp\,[\alpha(R_e-R)]\}, \quad (1)$$

where the potential minimum corresponds to the equilibrium distance R_e between atoms for a diatomic molecule. In what follows we set the dissociation energy of the diatomic molecule equal to $D=1$. We used a standard computer program to determine the energy of a cluster for a given configuration of its atoms, so that the total binding energy of the cluster atoms is defined as

$$E=-\sum_{i,k}\,U(r_{ik}), \quad (2)$$

where i,k label the atoms and r_{ik} the distances between them.

We must take into account that the interaction with non-nearest neighbors changes the distance between nearest neighbors to a value different from the equilibrium distance R_e for a diatomic molecule. In doing do we make use of the standard method[4] for the Morse interaction potential, writing the total energy of the cluster atoms Eq. (2) in the form

$$E=2\exp\,(\alpha R_e)F(\alpha a)-\exp\,(2\alpha R_e)F(2\alpha a), \quad (3)$$

where

$$F(\alpha a)=\sum_{i,k}\,\exp\,(-\alpha r_{ik}), \quad (4)$$

a is the distance between nearest neighbors (the basis for the given cluster), and the sum runs over all bonds. We then take the derivative of this function, which is

$$F'(\alpha a)=\frac{dF}{d(\alpha a)}=\sum_{i,k}\,r_{ik}\exp\,(-\alpha r_{ik}). \quad (5)$$

Once we have optimized the energy of the cluster E, we find the optimum distance a between nearest neighbors from the ratio

$$\exp\,(\alpha R_e)=g(\alpha a)=F'(\alpha a)/F'(2\alpha a), \quad (6)$$

so that the optimum binding energy of the cluster under study is

$$E=g(\alpha a)F(\alpha a)-\frac{1}{2}\,g^2(\alpha a)F(2\alpha a). \quad (7)$$

We then separate in the sum Eq. (4) the terms that involve nearest-neighbor F_{nn} and non-nearest neighbor interactions F_{nnn}, so that $F=F_{nn}+F_{nnn}$. This allows us to conveniently divide up the cluster energy in the standard way[13] into a nearest-neighbor interaction energy E_{nn}, a non-nearest neighbor interaction energy E_{nnn}, and a strain energy $E_{\rm str}$:

$$E=E_{nn}+E_{nnn}+E_{\rm str}. \quad (8)$$

Here

$$E_{nn}=2\exp\,(\alpha R_e)F_{nn}(\alpha a)-\exp\,(2\alpha R_e)F_{nn}(2\alpha a). \quad (9)$$

We define the interaction energy of non-nearest neighbors in the same way:

$$E_{nnn}=2\exp\,(\alpha R_e)F_{nnn}(\alpha a)-\exp\,(2\alpha R_e)F_{nnn}(2\alpha a). \quad (10)$$

In these expressions, the quantity a acts as the basis for the short-range interaction, which includes only the interaction between nearest neighbors. The strain energy arises from the change in the nearest-neighbor distance in the cluster from a to its optimal value.

Let us first discuss the case of short-range interactions between atoms in the cluster, when only the interaction between nearest neighbors is taken into account. Then for an *fcc* cluster we obtain

$$F_{nn}(a)=K\exp\,(-\alpha a), \quad F_{nnn}=0, \quad (11)$$

where K is the total number of bonds between nearest neighbors for the given cluster. If we take the value of this quantity from Ref. 17, we have $F'_{nn}=-F_{nn}$, so that optimization of the energy gives $g(\alpha a)=\exp(-\alpha a)$ and $a=R_e$ for nearest neighbors. Then we obtain

$$E=K. \quad (12)$$

The interaction energy of non-nearest neighbors E_{nnn} and the strain energy $E_{\rm str}$ both equal zero.

For an icosahedral cluster, there are two distances between nearest neighbors. Let us denote the distance between nearest neighbors belonging to adjacent layers by R, and the distance between nearest neighbors in the same layer by R_0; then $R=0.951R_0$. Introducing A, the number of bonds between nearest neighbors in adjacent layers (bond length R) and B, the number of bonds between nearest neighbors in the same layer (bond length R_0), we obtain

TABLE I. Cluster parameters (m is the number of filled layers of the icosahedral cluster, n the number of cluster atoms.

m	n	A	B	K	$\alpha_* R_e$
3	147	276	420	668	8.21
4	309	648	900	1507	6.60
5	561	1260	1650	2844	6.07
6	923	2172	2730	4809	5.54
7	1415	3444	4200	7527	4.96
8	2057	5136	6120	11118	4.44

$$F_{nn}=A\ \exp\ (-\alpha R)+B\ \exp\ (-\alpha R_0). \tag{13}$$

Optimizing the binding energy of the cluster atoms, we find for the optimum distance between nearest neighbors in neighboring layers[11,12]

$$R=R_e[1-0.047B/(0.904A+B)]. \tag{14}$$

If we limit ourselves only to interaction between nearest neighbors, this leads to the following expression for the total energy of an icosahedral cluster with Morse-potential interactions between atoms:[11,12]

$$E=A+B-0.0024\alpha^2 AB/(0.904A+B). \tag{15}$$

Equality of the icosahedral and *fcc* cluster energies for short-range interactions between atoms in the cluster corresponds to the following value of the parameter in the Morse interaction potential:

$$\alpha_*=20.4[(A+B-K)(0.904A+B)/(AB)]^{1/2}. \tag{16}$$

In Table I we compare the parameters of clusters whose number of atoms corresponds to filled layers of the icosahedral structure. Table II lists these parameters versus the number of atoms for icosahedral clusters in which 10 surface triangles are occupied above each filled layer. In these tables we list values of the parameter α_* corresponding to energy level crossing, which is determined by Eq. (16). This comparison identifies the range of values of parameter α in which a competition between the structures takes place.

Next, we include the interaction of non-nearest neighbors as a perturbation, i.e., we construct a perturbation theory for Morse-potential interactions between cluster atoms in which the small parameter is $\Delta=F_{nnn}(\alpha R_e)/F_{nn}(\alpha R_e)$. This perturbation theory is valid for $\alpha>4$. In Table III we list values of this small parameter for *fcc* (Δ_{fcc}) and icosahedral clusters ($\Delta_{\rm ico}$). Table III also lists values of the parameter $\gamma=-d\ \ln F_{nnn}(\alpha R_e)/d\ \ln(\alpha R_e)$. Since R_e/γ is the characteristic distance between cluster atoms, which determines the value of F_{nnn}, it follows that $\gamma>1$. For this reason, we may neglect the second term in Eq. (10) compared to the first. Actually, in the range of cluster sizes of interest $n=561-923$, the quantity $\exp(\alpha R_e)F_{nnn}(2\alpha R_e)/F_{nnn}(\alpha R_e)$ equals 0.11 when $\alpha R_e=4$, 0.074 when $\alpha R_e=5$, and 0.054 when $\alpha R_e=6$.

TABLE II. Parameters of icosahedral clusters having m filled layers and 10 occupied surface triangles above each filled layer (n is the number of cluster atoms).

m	n	A	B	K	$\alpha_* R_e$
4	216	442	603	1024	5.74
5	420	928	1203	2094	5.31
6	724	1684	2103	3730	4.93
7	1148	2770	3363	6042	4.89
8	1712	4246	5043	9181	4.32

TABLE III. Non-nearest neighbor interaction parameters for clusters with the Morse-potential interaction between atoms.

αR_e	γ_{fcc}	Δ_{fcc}	$\gamma_{\rm ico}$	$\Delta_{\rm ico}$
4	1.68	0.21	1.72	0.23
5	1.61	0.11	1.64	0.12
6	1.56	0.061	1.60	0.071
7	1.52	0.036	1.56	0.042
8	1.50	0.021	1.53	0.026
10	1.46	0.0082	1.50	0.010

Thus, within the framework of this perturbation theory, we obtain for binding energies of atoms of an *fcc* cluster

$$E=2K\ \exp\ [\alpha(R_e-a)]-K\ \exp\ [2\alpha(R_e-a)] +2F_{nnn}(\alpha a)\exp\ (\alpha R_e). \tag{17}$$

Optimizing the cluster energy with respect to distance between nearest neighbors a, we obtain for this distance

$$\alpha(R_e-a)=\gamma F_{nnn}(\alpha R_e)/F_{nn}(\alpha R_e)=\gamma\Delta. \tag{18}$$

From this, we find for the terms in Eq. (8)

$$E_{nn}=K,\quad E_{nnn}=2K\Delta,\quad E_{\rm str}=\gamma^2\Delta^2 K. \tag{19}$$

In a similar way we obtain for the total binding energy of atoms in an icosahedral cluster

$$E=A+B-0.0024\alpha^2 AB/(0.904A+B)-(A+1.106B) \times(R'-R)^2+2F_{nnn}(\alpha R)\exp\ (\alpha R_e), \tag{20}$$

where the distance R' is given by Eq. (14): $R'=R_e[1-0.047B/(0.904A+B)]$. After optimizing the binding energy of the cluster atoms for a given value of the parameter αR_e according to the parameter R, we obtain for the individual terms of Eq. (8)

$$E_{nn}=A+B-0.0024\alpha^2 AB/(0.904A+B),$$

$$E_{nnn}=2F_{nnn}(\alpha R')\exp(\alpha R_e),$$

$$E_{\rm str}=(\gamma\alpha)^2 F_{nnn}^2(\alpha R')\exp\ (2\alpha R_e)/(A+1.106B). \tag{21}$$

As in the case of *fcc* clusters, it follows from these expressions that $E_{nnn}/E_{nn}\sim\Delta$, $E_{\rm str}/E_{nn}\sim\Delta^2$.

Note that for a given value of the parameter αR_e the specific quantities F_{nn}/n and F_{nnn}/n are smooth functions of the number of atoms n in the cluster. For example, in the interval of cluster sizes we are interested in the quantity K changes from 2844 for $n=561$ to 4809 for $n=923$. Accordingly, the quantity K/n changes from 5.07 to 5.21, i.e., by a few percent. In contrast to the smooth function that describes this quantity, the number of bonds between nearest neighbors as a function of n is discontinuous at magic numbers of atoms in the cluster; however, the difference of these functions even for the magic numbers of atoms does not exceed

TABLE IV. Parameters for competition between *fcc* and icosahedral structures.

n	A	B	K	$\alpha_* R_e$
561	1260(ic)	1650	2844(nc)	7.1
606	1395(fcc)	1740	3093(nc)	6.8
636	1485(fcc)	1800	3255(nc)	6.4
681	1620(fcc)	1890	3490(c)	6.2
688	1592(ic)	1996	3529(c)	6.7
724	1677(ic)	2103	3730(c)	6.7
760	1774(ic)	2210	3934(c)	6.6
817	1916(ic)	2385	4234(c)	6.8
874	2058(ic)	2564	4575(c)	6.5
923	2172(ic)	2730	4809(c)	7.1

20, i.e., is relatively small compared to the total number of bonds. The function F_{nnn}/n is a smoother function of the number of atoms in the cluster, since it describes the interaction of non-nearest neighbors. Thus, in calculating the binding energy of atoms in a cluster we may assume that the specific energetic parameters do not depend on n within a certain range of n. However, this assumption can lead to errors in analyzing the competition between *fcc* and icosahedral structures.

Actually, it follows from the data of Tables I, II that the numbers of bonds between nearest neighbors are close for optimal configurations of atoms in *fcc* and icosahedral structures, so that their relative difference is a few percent over a wide range of cluster parameters. From Table III it is clear that the non-nearest neighbor interaction energies are also close. Table IV summarizes the results of a comparison of clusters with different structures, for each of which the optimum configuration of atoms has been chosen. While the core of the cluster can have only an icosahedral structure, its (outermost) filling layer can have either *fcc* or icosahedral structure. The structure of this surface layer is indicated next to the number of bonds between nearest neighbors by the letters *ic* or *fcc* respectively. An *fcc* structure is preferred for the surface layer of an icosahedral cluster as long as no more than eight surface triangles of the cluster are occupied. For larger numbers of occupied triangles a surface layer with icosahedral structure becomes energetically more favorable.

As for *fcc* clusters, there is also a competition between structures with and without a central core atom. As the number atoms in the cluster increases, these structures are found to alternate. Within the range of cluster sizes under discussion here, the structure with a central core atom turns out to be energetically favorable for the most part. The preferred core structure of the *fcc* clusters is indicated in Table IV by the letters *c* or *nc* in brackets next to the number of bonds K between nearest neighbors of the cluster, denoting the presence or absence of a central atom in the cluster core.

Table IV lists values of the parameter $\alpha_* R_e$ at which the binding energy of a cluster consisting of this number of atoms becomes the same for *fcc* and icosahedral cluster structures for optimum configurations of atoms. Equations (19), (20) were used for the *fcc* and icosahedral cluster energies. The parameters for the optimum configuration of an icosahedral cluster with short-range interactions between the atoms were taken from Refs. 11, 12, while the parameters for the optimum configuration of an *fcc* cluster are from Refs. 14, 17. The parameters for long-range interactions between cluster atoms, including the interaction between non-nearest neighbors, were calculated for a Morse-potential interaction by numerical methods using Eq. (4). Note that for the interval of cluster sizes $n=561-923$ being analyzed, the values of the parameter $\alpha_* R_e$ of interest in the range in which these cluster structures compete are concentrated in the range $\alpha_* R_e = 6-7$, so that the interaction between nearest neighbors gives the primary contribution to the total binding energy of the cluster atoms. This justifies our method, in which we choose the optimum configurations of atoms for short-range atomic interactions. Nevertheless, the interactions of non-nearest neighbors are important in analyzing the competition between structures.

Actually, the data of Tables I, II, IV imply that the numbers of bonds between nearest neighbors are similar for the *fcc* and icosahedral clusters. Hence, although the non-nearest neighbor interaction potential in the cluster is small compared to the interaction between nearest neighbors, it can turn out to be comparable to the difference between the nearest neighbor interaction potentials for these structures. In fact, it follows from Tables I, II, IV that the difference $A+B-K$ entering into Eq. (16) is a few percent of $A+B$. This is the same order of magnitude as the interaction potential between non-nearest neighbors in the range of parameter values $\alpha R_e = 6-7$, where these structures compete (see Table III). For the icosahedral structure the interaction potential between non-nearest neighbors is larger than it is for the *fcc* structure, so that this interaction increases the energy level crossing parameter $\alpha_* R_e$ above its value when the interaction includes nearest neighbors only. Note that in the energy level crossing range the strain energy, according to Eqs. (19), (21), is proportional to Δ^2, and hence is unimportant.

The quantity $A+B-K$ is sensitive to the structure of the clusters, exhibiting local maxima at magic numbers of the icosahedral cluster and minima at magic numbers of the *fcc* cluster. Consequently, in a range of cluster sizes that includes magic numbers the quantity $\alpha_* R_e$ will be an irregular function of n. The non-nearest neighbor interaction makes this function smoother, since this interaction is insensitive to the magic numbers.

Note that the popular Lennard–Jones potential corresponds to a Morse potential with the parameter $\alpha R_e = 6$. Then our analysis implies that for atoms interacting via the Lennard–Jones potential these cluster structures will compete for $n > 1000$, which corresponds to Ref. 11.

3. CONCLUSION

Thus, we have shown that the competition between icosahedral and *fcc* cluster structures for optimal configurations of atoms, equal numbers of atoms, and pairwise interaction potentials between these atoms, is conveniently analyzed in terms of the parameter αR_e of the Morse-potential for which the energies of these clusters coincide.[13] We have calculated these cluster energies numerically for clusters containing $n=561-923$ atoms, which corresponds to filling

of the sixth layer of an icosahedral cluster. It follows from these calculations and our analysis that our energy level crossing parameter $\alpha_* R_e$ is determined by interactions between both nearest neighbors and non-nearest neighbors in the cluster. In the range of cluster sizes of interest, this parameter varies in the range $\alpha_* R_e = 6 - 7$ and is a nonmonotonic function of the number of atoms in the cluster. Within this range of values of the parameter αR_e the primary contribution to the cluster energies of both structures comes from the interaction between nearest neighbors, so that the interaction of non-nearest neighbors is a few percent of the total cluster energy. Nevertheless, since the number of bonds between nearest neighbors turns out to be close for the optimal configurations of atoms of both structures, it is important to include the interaction between non-nearest neighbors when determining the parameter $\alpha_* R_e$ at which the energies of these structures coincide. Interactions between non-nearest neighbors increase the value of $\alpha_* R_e$ and make the dependence of this parameter on the number of atoms in the cluster smoother.

This work was partially supported by the Russian Fund for Fundamental Research (Grant 96-02-16016).

[1] M. J. Buerger, *Elementary Crystallography* (John Wiley, New York, 1963).
[2] R. W. G. Wyckoff, *Crystal Structure* (Interscience, New York, 1963).
[3] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehardt, and Winston, New York, 1976).
[4] C. Kittel, *Solid State Physics*, 6th edition (John Wiley, New York, 1986).
[5] A. L. Mackay, Acta Crystallogr. **15**, 916 (1962).
[6] J. Farges, M. F. deFeraudi, B. Raoult, and G. Torchet, J. Chem. Phys. **78**, 5067 (1983).
[7] J. Farges, M. F. deFeraudi, B. Raoult, and G. Torchet, J. Chem. Phys. **83**, 3491 (1986).
[8] J. W. Lee and G. D. Stein, J. Chem. Phys. **91**, 2450 (1987).
[9] B. W. van de Waal, J. Chem. Phys. **90**, 3407 (1989).
[10] J. Xie, J. A. Northby, D. L. Freeman, and J. P. Doll, J. Chem. Phys. **91**, 612 (1989).
[11] B. M. Smirnov, Chem. Phys. Lett. **232**, 395 (1995).
[12] B. M. Smirnov, Usp. Fiz. Nauk **164**, 665 (1994) [Phys. Usp. **37**, 1079 (1994)].
[13] J. P. K. Doye, D. J. Wales, and R. S. Berry, J. Chem. Phys. **103**, 4234 (1995).
[14] B. M. Smirnov, Phys. Scr. **52**, 710 (1995).
[15] B. M. Smirnov, High Temp. **33**, 694 (1995) [in Russian].
[16] J. A. Northby, J. Chem. Phys. **87**, 6166 (1987).
[17] B. M. Smirnov, Zh. Éksp. Teor. Fiz. **107**, 2080 (1995) [JETP **80**, 1151 (1995)].

Translated by Frank J. Crowne

Stabilization of electron–hole liquid in uniaxially strained germanium in a strong magnetic field

A. V. Chernenko and V. B. Timofeev

Institute of Solid State Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia
(Submitted 14 February 1997)
Zh. Éksp. Teor. Fiz. **112**, 1091–1105 (September 1997)

Luminescence spectra of uniaxially and uniformly strained high-purity germanium crystals at liquid-helium temperatures in a magnetic field of up to 14 T have been investigated. In strongly strained Ge crystals, a new line has been detected on the low-energy side of the exciton line in magnetic fields higher than 4 T. Studies of this line's characteristics as functions of pressure, temperature, and magnetic field have led us to conclude that its presence is due to recombination of electron–hole pairs in an electron–hole liquid. The experimental data suggest that the metallic electron–hole liquid is stabilized in a strong magnetic field. By approximating the shape of the newly detected line using the model of metallic electron–hole liquid, we have obtained the electron–hole liquid density $n_{\mathrm{EHL}}(B)$ and Fermi energies $E_{Fe,h}$ of electrons and holes. The liquid binding energy ϕ as a function of magnetic field has been estimated. © *1997 American Institute of Physics.* [S1063-7761(97)02109-4]

1. INTRODUCTION

Uniaxially and uniformly strained germanium crystals present a unique opportunity for studying properties of the electron–hole (EH) system in a magnetic field. Under a strong uniform strain, the valence band spectrum in Ge is simplified considerably due to the splitting between light and heavy hole levels, which were initially degenerate at the Brillouin zone center. This valence band splitting leads to a lower cyclotron mass of holes. In Ge strained along an axis close to [100] (we will denote this material as Ge[$\sim$100]), the ratio $\gamma=\hbar\omega_c/2\mathrm{Ry}$ ($\hbar\omega_c$ is the cyclotron energy and Ry is the exciton binding energy at zero magnetic field), equals 4.8 at a field of 14 T. Moreover, in Ge[$\sim$100] the valley–orbit degeneracy in the conductance band is lifted, and germanium becomes a one-valley semiconductor. In magnetic fields higher than 4 T and at liquid-helium temperatures, when the condition $\mu_B|g_{e,h}|B>k_BT$ is satisfied in Ge[$\sim$100] (here μ_B is the Bohr magneton and $g_{e,h}$ are g-factors of electrons and holes), high-density gas of spin-aligned excitons, $n_{\mathrm{exc}}\sim10^{15}\,\mathrm{cm}^{-3}$, can be created. In this system produced by a uniform and uniaxial strain, one can expect properties predicted for dense EH ''matter'' in the limit of strong magnetic field, $\gamma\gg1$. Note that in unstrained Ge the parameter $\gamma<1$ for realistic magnetic fields.

A magnetic field which is so strong that its influence is comparable to or greater than the Coulomb energy considerably changes properties of the EH system. The transverse exciton dimension in a strong magnetic field is determined not by the Bohr radius a_x, but by the magnetic length $\lambda_m=\sqrt{\hbar c/eB}\ll a_x$. Since the exciton dimension along the magnetic field decreases significantly more slowly ($\sim\ln^{-1}B$),[1] it looks like a ''needle'' aligned with the magnetic field. It has a nonzero quadrupole moment. The exciton binding energy E_0 increases with magnetic field as $\ln^2 B$ (provided that $\ln B\gg1$).

In theoretical studies of the EH system in the limit $\gamma\to\infty$, new bound states were predicted. Godiak *et al.*[2] asserted that a triplet excitonic molecule (biexciton) composed of excitons bound by van der Waals forces and quadrupole–quadrupole coupling should be stabilized in a strong magnetic field. In addition to the biexciton, they predicted emergence of a multiexciton cluster similar to a molecule. A strong magnetic field can also lead to formation of a liquid dielectric phase with coupling between electrons and holes similar to the Cooper pairing of electrons in superconductors.[3] Keldysh and Onishchenko[4] and Chui[5] predicted stabilization of a metallic electron–hole liquid and an increase in its binding energy and density in a strong magnetic field.

Earlier[6] we reported that in an ultrapure Ge strained along an axis close to [100] in a magnetic field higher than 4 T, a new X line appeared on the low-energy side of the exciton line, and its width and intensity grew with the magnetic field. In connection with the predictions discussed above, the nature of this line is quite interesting. In the present work, we have investigated properties of the X line and tried to explain how a high-density excitonic system in an indirect semiconductor with nondegenerate bands behaves in a strong magnetic field.

2. EXPERIMENTAL TECHNIQUES

For our measurements, we used Ge single crystals with residual concentrations of electrically active impurities $|N_d-N_a|\leqslant10^{11}\ \mathrm{cm}^{-3}$ and $|N_d-N_a|\leqslant10^{12}\ \mathrm{cm}^{-3}$. The samples were cut down to rectangular parallelepipeds with dimensions of $2.5\times2.5\times10$ mm. In the Ge[$\sim$100] samples, the longer edge made an angle of 5° with the [100] axis to within 1°. In Ge[111] samples the longer edge was aligned with the [111] axis to within 2°. In order to reduce surface recombination, the samples were treated with SR-4A polish-

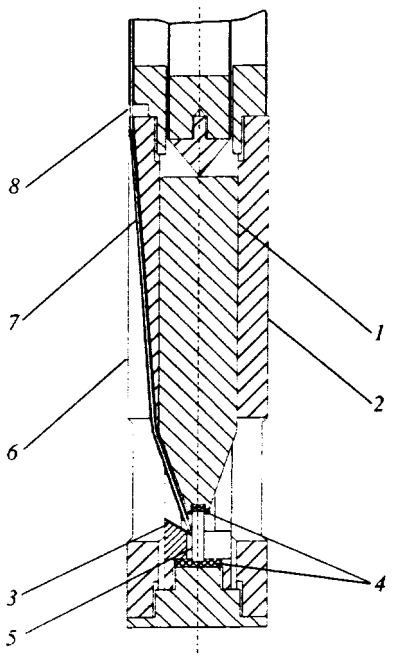

FIG. 1. Lower section of the experimental apparatus used in measurements of luminescence spectra of uniaxially strained Ge crystals immersed in liquid helium: *1*) piston; *2*) cylindrical casing; *3*) mirror; *4*) spacers (soldering alloy); *5*) sample; *6*) slot for an optical fiber in the casing; *7*) thin tube; *8*) optical fiber.

ing solution. Luminescence in a magnetic field of up to 14 T and strong uniaxial and uniform strain was measured using a specially designed apparatus.

The head of this apparatus, immersed in liquid helium, is schematically shown in Fig. 1. The main components are piston *1* tightly fitted to casing *2*. The piston could be driven in the casing by a force larger than 10 N. In order to produce uniform strain in the sample, one or two spacers *4* made from a Pb–Sn soldering alloy with a thickness of about 1 mm were inserted between the sample and compressing planes. Laser light was fed to the sample through optical fiber *8*, and luminescence radiation was collected by the same optical fiber. The fiber was conducted close to the sample *5* via bent stainless steel tube *7*. The design of our apparatus is described in detail elsewhere.[7]

Luminescence radiation was dispersed by a grating monochromator with a dispersion of 26 Å/mm in the spectral band studied. Nonequilibrium carriers in the sample were excited by a cw Nd:YAG laser emitting up to 3 W at wavelength $\lambda = 1.064$ μm.

The optical signal was detected by a cooled Ge(Au) light detector with a noise-equivalent power of $\sim 10^{-14}$ W/Hz$^{1/2}$ in the investigated spectral band. Its output was fed to a lock-in amplifier. The magnetic field and strain axis were parallel, and luminescence emitted perpendicular to the magnetic field was detected (Voigt configuration).

3. EXPERIMENTAL RESULTS AND DISCUSSION

Strain applied in the $[\sim 100]$ direction is the most efficient means of splitting the levels of light holes and heavy holes and reducing the effect of band nonparabolicity on the density of states in the valence band. This strain splits the four-fold degenerate level at the Brillouin zone center into two doubly degenerate levels. The coupling between the split levels leads to a nonparabolicity in the hole spectrum. This coupling decreases with the strain, and the effect of nonparabolicity on the density of states tends to zero.[8]

The uniaxial strain in the sample was estimated using the energy shift of the exciton line. The gap width E_g as a function of strain is described as follows:

$$\Delta E_g = M \mathrm{Tr}\ \varepsilon + \Xi_u \varepsilon^{0\min} - E_\varepsilon .$$

Here $\Xi_u = 16.2 \pm 0.4$ eV, $M = \Xi_d + (1/3)\Xi_u - a = -2 \pm 0.5$ eV are deformation potentials,[9] E_ε is half of the valence band splitting, $E_\varepsilon / P = (1/2)\Delta E_v / P = b(S_{11} - S_{12}) = 2.5 \pm 0.1$ meV/kbar ($S_{11} = 0.96$ kbar^{-1} and $S_{12} = -0.26$ kbar^{-1} are ductility coefficients, $b = 2.2 \pm 0.1$ eV,[10] ε_{ij} is the strain tensor, $\varepsilon^{0\min} = a_i\, \varepsilon_{ij}\, a_j - (1/3)\mathrm{Tr}\ \varepsilon$ is the shift of the lowest conductance band valley with respect to the mean position of the other valleys, and a_i is the unit vector in the direction of this valley.[11] For Ge$[\sim 100]$, $\Delta E_g / P \approx -2.7 \pm 0.2$ meV/kbar. The change in the exciton binding energy Ry due to deformation was also taken into account. In unstrained Ge we have Ry=4.15 meV, whereas in Ge strongly strained along the $[\sim 100]$ axis Ry=2.8 meV.[12]

In Ge strained along the [111] and $[\sim 100]$ axes, each of the split levels corresponds to a definite hole spin projection $s_z = \pm 1/2, \pm 3/2$. The lowest hole level has the spin projection $\pm 1/2$. The reduction in spin degeneracy in both the valence and conductance bands results in a higher kinetic energy per electron–hole pair, and lower binding energy and density of states of the electron–hole liquid (EHL).[10] This takes place at pressures P such that $\Delta E_{c,v} > E_{Fe,h}$. In Ge$[\sim 100]$ the critical pressure is less than 1 kbar. Under higher pressure, the hole kinetic energy increases, owing to the larger splitting between the valence band levels and the smaller effect of nonparabolicity on the hole density of states. In uniaxially strained Ge$[\sim 100]$ at liquid-helium temperatures and $P > 1.5$ kbar, the EHL phase becomes unstable.[13]

At $T \approx 2$ K, the luminescence spectrum of unstrained germanium contains only the line corresponding to EHL. In this case, the exciton gas density near the condensation threshold is about 10^{12} cm^{-3}. The exciton luminescence line is not observable, owing to the low excitation density. When uniaxial pressure is gradually applied, exciton and biexciton lines emerge in the luminescence spectrum, and their intensities rapidly increase. The line of EHL emission approaches to the exciton line as the pressure increases. It becomes narrower, and its intensity drops. At $P \sim 1.5$ kbar, the EHL line disappears from the spectrum. At free exciton densities up to the maximum value $n_{\mathrm{FE}} \cong 2 \cdot 10^{15}$ cm^{-3} realized in our experiments, the EHL spectral line did not show up again in the absence of a magnetic field. The component of the Ge$[\sim 100]$ luminescence spectrum with emission of an LA-phonon at $P = 1.5$ kbar and zero magnetic field is shown in Fig. 2. Below, we discuss only the LA-components of the luminescence spectra.

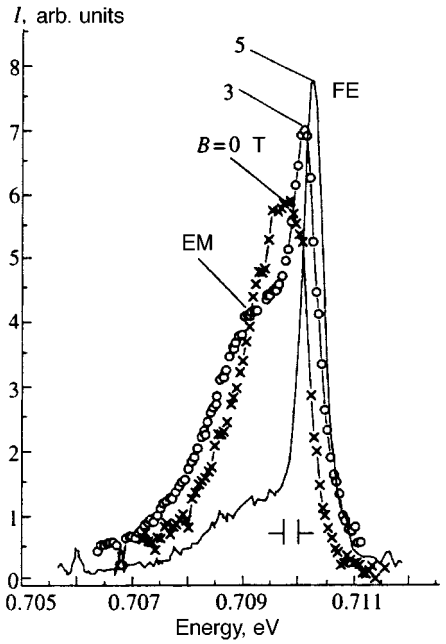

FIG. 2. Spectra of Ge[~100] luminescence with emission of LA-phonons in magnetic field of up to 5 T at $T=1.9$ K, $P=1.5$ kbar, $W=50$ W/cm^2. Luminescence lines of free excitons (FE) and excitonic molecule (biexciton) (EM) are shown. At $B=0$ T (this spectrum is shown by empty circles) the FE and EM lines are not resolved.

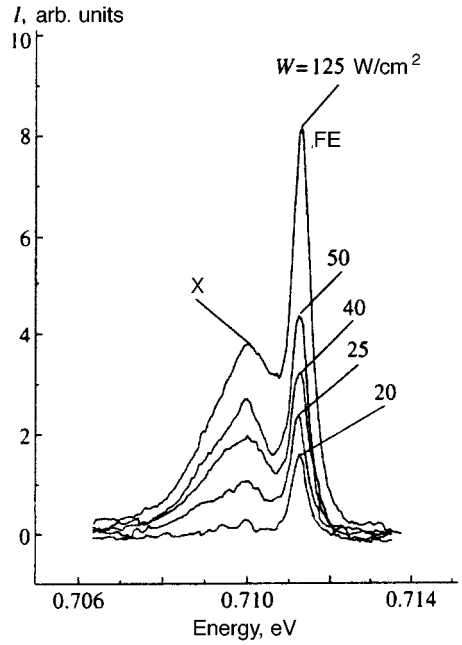

FIG. 3. LA-phonon components in Ge[~100] luminescence spectra at constant temperature $T=1.9$ K, $P=1.5$ kbar, $H=9$ T, and different pump power densities W.

The spectra in Fig. 2 contain the free exciton line FE and the excitonic molecule (biexciton) line EM. When magnetic field is applied, the excitonic molecule is destroyed. At $B>4$ T, the EM line disappears from the spectrum (Fig. 2).[14] Given the equality between the chemical potentials of excitons and biexcitons, $2\mu_{FE}=\mu_{EM}$, the ratio between the total intensities I_{FE}/I_{EM} of the FE and EM lines yields the most accurate estimate of the exciton gas density n_{FE}. The maximum density observed in our experiments was $2.2\cdot 10^{15}$ cm^{-3} (the parameter $r_s=3/4\pi a_x n^{1/3}\approx 3$). The laser power density in this case was $W\approx 120$ W/cm^2. The strain nonuniformity was derived from the broadening of the FE line under pressure P. At $P=2.8$ kbar the nonuniformity ΔP was of order 2%.

The exciton gas temperature was derived from the exciton line FWHM. The profile of the exciton luminescence line in strained Ge is described by the expression $\sqrt{E}\exp(-E/k_BT)$. Its FWHM is $1.8k_BT$. The exciton gas temperature under excitation by a Nd:YAG laser and pumping of helium vapor was ~1.9 K, which was 0.3 to 0.5 K higher than the helium bath temperature.

When the magnetic field was higher than 4 T, a ''pedestal'' was observed on the low-energy side of the exciton line, i.e., a flat and wide line without a clearly defined peak.[6] The width of this line was notably larger than that of the FE line. Its width and amplitude increased with magnetic field, and the pedestal transformed to a clearly defined X line. The X line maximum shifted to the low-energy side with respect to the exciton line. As the pumping density generated by an Ar^+ laser increased, the X line intensity remained constant, whereas the FE line intensity rapidly grew. In order to elucidate the nature of this line, we investigated its parameters in the present work as functions of magnetic field, pressure, pumping density, and temperature.

Figure 3 shows luminescence spectra of Ge[~100] at $P=1.5$ kbar and various pumping densities. In the present work, we excited our samples using a Nd:YAG laser, which enabled us to reduce overheating of the EH system considerably. The exciton temperature in pumped helium under excitation by the Ar^+ laser was $T\approx 3$ K,[6] whereas with the Nd:YAG laser it was ~1.9 K. The maximum exciton density generated by the Nd:YAG laser was five times greater than in the case of the Ar^+ laser.

Figure 3 indicates that in the pump power range $W=20-50$ W/cm^2, the X line intensity increases more rapidly than that of the FE line. At $W<15$ W/cm^2, the X line is not observed in the spectrum. At low pump powers, the X line intensity strongly fluctuates throughout its spectral width (Figs. 4, 5a, and 6). This behavior indicates that the system is near a phase transition, and the X line corresponds to luminescence of a collective electron–hole state whose binding energy is higher than the exciton binding energy.

For pump powers higher than 50 W/cm^2, I_{FE}, on the contrary, increases faster than I_X. This is an indication that at high pump powers, overheating of the dense EH system by laser light becomes significant. The effect of the phonon wind also must be taken into account. The observed independence of the X line intensity from the pump power reported previously[6] was, most probably, caused by laser overheating of the EH system.

Luminescence spectra of Ge[~100] at $T=1.9$ K recorded at various pressures (Figs. 4, 5, and 6) demonstrate a strong dependence of the X-line width and intensity on the strain.

Under high strain, the peak of the X line shifts with

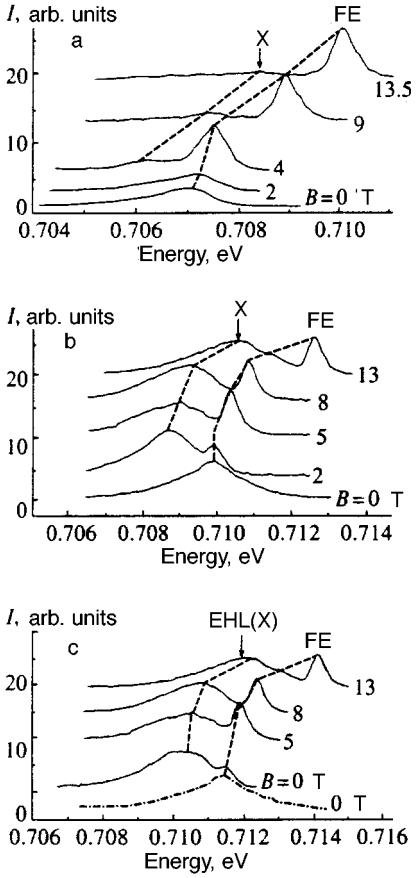

FIG. 4. LA-phonon components of Ge[∼100] luminescence spectra in magnetic fields of up to 14 T at $T=1.9$ K, various pressures and pump power densities: a) 2.8 kbar, $W=120$ W/cm^2; b) 1.6 kbar, $W=80$ W/cm^2; c) 1.2 kbar, $W=80$ W.cm^2. In graph c, the lower spectrum ($B=0$ T) shown by a dot-and-dash line was recorded at $T=4.2$ K. The peaks of the FE and EHL(X) lines are connected by the dashed lines.

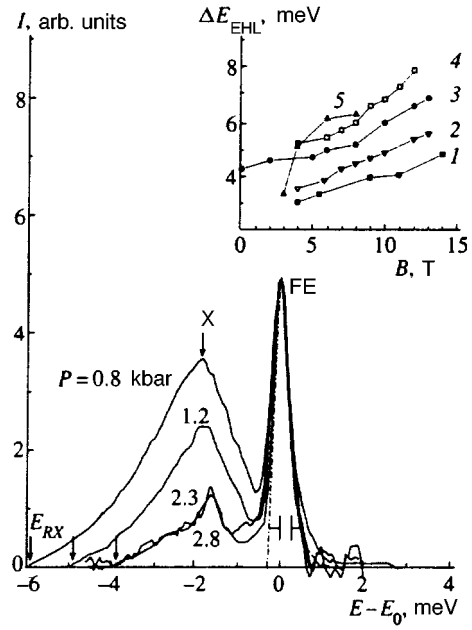

FIG. 6. LA-phonon components of Ge[∼100] luminescence spectra recorded at a constant magnetic field $B=12$ T, constant temperature $T=1.9$ K, various pressures, and similar pump powers. The dot-and-dash line shows an approximation of the FE line that allows for the monochromator instrumental function. The arrows mark the low-energy edge of the EHL line. We define the position of this edge as a point where the luminescence intensity is a factor of 25 lower than at the line peak. The high-energy edge of the EHL line is assumed to coincide with the low-energy edge of the FE line. The abscissa corresponds to the difference $E-E_0$. All spectra are normalized to the FE line peak. The insert shows the width of the EHL line, ΔE_{EHL}, in strained germanium as a function of magnetic field at a constant pump power determined as described above. Curves *1–4* correspond to Ge[∼100]: *1*) 2.8 kbar; *2*) 1.6 kbar; *3*) 1.2 kbar; *4*) Ge[111] at 1.1 kbar; *5*) 0.5 kbar.

magnetic field to the high-energy side, and its intensity and width increase with B. An increase in the uniaxial pressure P leads to a notable decrease in the X line intensity relative to FE. The spacing between the peaks of X and FE lines also decreases. As the strain is lifted, both the intensity and width of the X line grow rapidly, and the line is transformed to an EHL line.

Under conditions of the low strain, $P=1.2$ kbar (Fig. 4c) and $P=0.5$ kbar (Fig. 5a), when the EHL line is easily identified in the spectra, its behavior is similar to that of the X line in magnetic field. When the pump power and magnetic field are varied, the EHL line in Ge[∼100] varies identically for P ranging between 0.5 and ∼1.5 kbar. The liquid line in Ge[111] behaves similarly under a pressure of up to the maximum 2.8 kbar attained in experiments. A luminescence spectrum of Ge[111] for $P=1.1$ kbar is given in Fig. 5b.

Note that the total intensity of the liquid line rapidly increases with B in the range between 4 and 8 T, whereas for $B>8$ T it is essentially constant. The intensity of the FE line

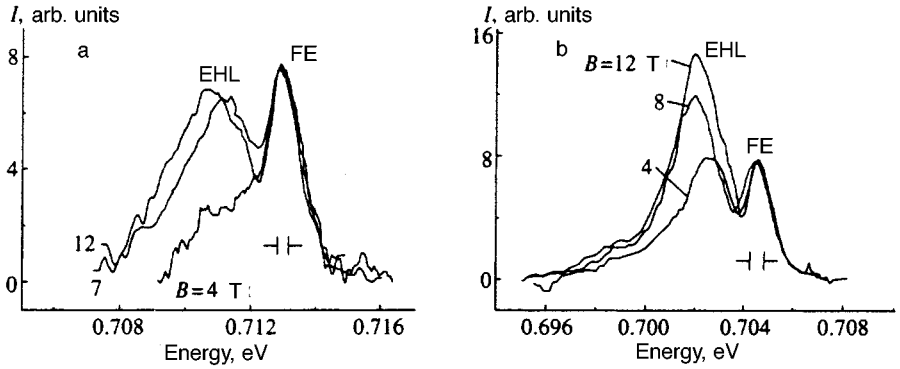

FIG. 5. LA-phonon components of luminescence spectra: a) Ge[∼100] at pressure $P=0.5$ kbar, $T=3$ K. The pump power is generated by an Ar^+ laser; b) Ge[111], $P=1.1$ kbar, $T=1.9$ K. The pumping is due to a Nd:YAG laser.

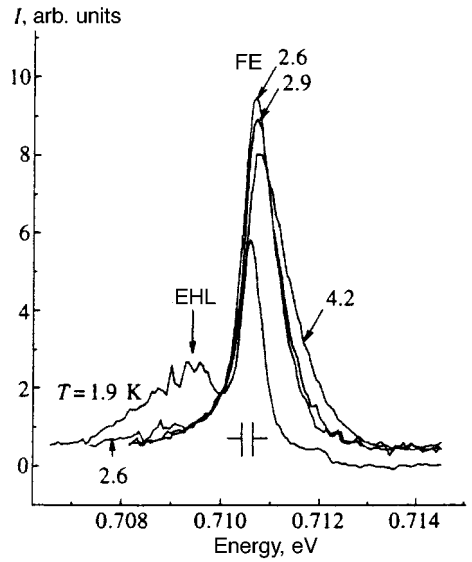

FIG. 7. Ge[$\sim$100] spectra with emission of an LA-phonon at constant pressure $P=1.5$ kbar in a magnetic field $B=7$ T, a pump power density $W=35$ W/cm^2, and various temperatures.

also increases in fields ranging up to 8 T, while in the range of 8–14 T it drops by 15–20%.

The width of the X(EHL) line versus magnetic field at various pressures is shown in Fig. 5. The apparent width of the X line (ΔE_{EHL}) increases with B at all applied pressures in the range 0.5–2.8 kbar (Fig. 4).

Thus, investigations of the X line in magnetic field versus temperature and pressure clearly indicate that it is due to EHL luminescence.

In order to determine the EHL binding energy ϕ as a function of B, we measured the temperature T_0 at which the EHL line vanishes under conditions of constant pump power in the magnetic field range 4–14 T. It turned out that for $B=4-14$ T, the critical temperature T_0 lies in the interval 2.6–2.9 K. Luminescence spectral of Ge[$\sim$100] at $P=1.5$ kbar in the temperature range 1.9–4.2 K are given in Fig. 7.

We derived ϕ using the expression for the density of an excitonic gas in equilibrium with the liquid:

$$n_{\mathrm{FE}}=\left(\frac{M_{\mathrm{FE}}k_BT}{2\pi\hbar^2}\right)^{3/2}\exp\left(-\frac{\phi'}{k_BT}\right), \tag{1}$$

where $\phi'=\phi-2\sigma/n_{\mathrm{EHL}}R$, σ is the surface tension, R is the liquid drop radius, n_{FE} is the excitonic gas density, M_{FE} is the exciton density-of-states mass, which is approximately $0.35m_0$, and m_0 is the free electron mass. The term responsible for the surface tension is negligible at temperatures higher than 1 K, so we neglected this effect. Using Eq. (1), we have obtained $\phi\approx 0.35\pm 0.05$ meV.

It is well known that ϕ derived using Eq. (1) is smaller than that obtained by processing spectroscopic data.[15] The reason for this is that Eq. (1) does not take into account an excitonic flow associated with the need to compensate for carriers lost to recombination in liquid drops. As a result of this exciton flow to the drops, the excitonic gas density near the drop surface is higher than the equilibrium value. In the central section of the phase diagram, the exciton density is 20 to 30% higher than the equilibrium value. As the system moves along the phase equilibrium curve toward lower n_{FE}, this deviation from the equilibrium density rises, and n_{FE} can be several times the equilibrium value.[15]

Although measurements of T_0 indicate that ϕ is in fact constant with B, the drop in the total intensity of the excitonic line is an indication of an increase in ϕ, which can be naturally attributed to a drop in exciton density.

4. ANALYSIS OF ELECTRON–HOLE LIQUID LUMINESCENCE LINE PROFILE

To obtain more accurate estimates of the liquid binding energy and determine its density, we analyzed the shape of its luminescence line. In our analysis, we assumed that the electron–hole liquid was metallic. We analyzed the spectra recorded at the maximum pressure $P=2.8$ kbar, when the metallic liquid had the highest binding energy and density. Another convenient feature of the high-strain limit is that the hole effective mass can be considered independent of the quasimomentum.

In calculating the EHL line profile, we largely used the technique suggested by Störmer and Martin,[16] who successfully described the EHL line profile in unstrained germanium in magnetic fields of up to 19 T.

The spectral density of the LA line is

$$I(h\nu,\mu(n_{\mathrm{EHL}},\Gamma_0))=A\int_0^\infty D_e(E_e)f(E_e,E_{Fe},T)\times D_h(E_h)f(E_h,E_{Fh},T)\,dE_e, \tag{2}$$

where $D_e(E_e)$, $D_h(E_h)$, $f(E_e,E_{Fe})$, and $f(E_h,E_{Fh})$ are the density of states and Fermi functions of electrons and holes, respectively, and E_{Fe} and E_{Fh} are the respective Fermi energies. The energies of recombining electrons and holes are related by $E_h=E_e+h\nu-E_g^{\mathrm{LA}}$, where E_g^{LA} is the band gap minus the LA-phonon energy (27.7 meV). We assume that the matrix element of the recombination transition is included in the constant A and is independent of the quasimomenta of both electrons and holes. The density of states of electrons and holes is

$$D=\frac{1}{4\pi^2\lambda_m^2}\left(\frac{2m^*}{\hbar^2}\right)^{1/2}\times\sum_n\left[E-\left(n+\frac{1}{2}\right)\hbar\omega_c\pm\frac{g}{2}\mu_BB\right]^{-1/2}.$$

The electron and hole cyclotron masses in the expression for ω_c are $m_{ce}=0.135m_0$, $m_{ch}=m_{th}=0.109m_0$. The translational masses m^* along the magnetic field are $m_h^*=m_{lh}=0.04m_0$, $m_e^*=m_{el}\cos^2\theta+m_{et}\sin^2\theta=0.584m_0$. Here $m_{hl}=(A+B)^{-1}m_0$ is the hole mass along the strain axis, and $m_{ht}=(A-B/2)^{-1}m_0$ is the hole mass in the transverse direction, $A=13.38$ and $B=8.48$ are the Ge valence band parameters,[8] m_{el} and m_{et} are the longitudinal and transverse electron masses, and θ is the angle between the strain axis ([$\sim$100]) and the electron ellipsoid rotation axis.

To allow for the finite lifetime of electrons and holes in EHL, we introduced broadening of energy levels[16]:

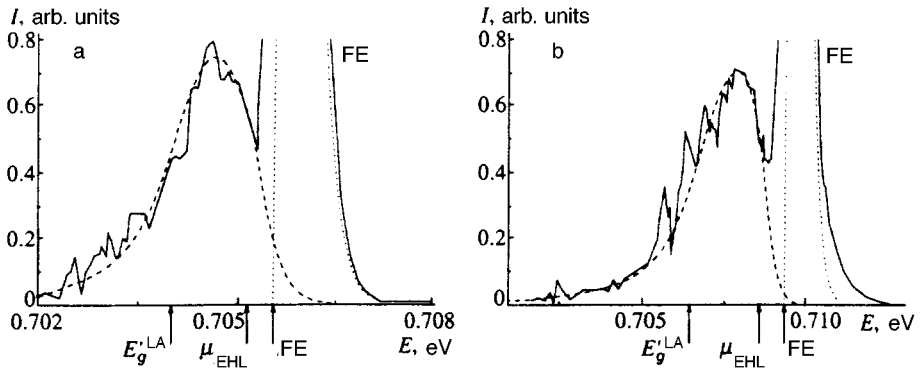

FIG. 8. Approximations of the EHL LA-phonon-assisted luminescence lines for $P=2.8$ kbar at a) $B=5.5$ T and b) $B=13.5$ T.

$$D^*_{e,h}(E,B)=\int_0^\infty L(E,E')D(E',B)\,dE',$$

$$L(E,E')=\begin{cases}L_0(E,E'), & E'<E\\ \delta(E-E'), & E'>E_f,\end{cases}$$

$$L_0(E,E')=\frac{1}{2\pi}\,\frac{\Gamma(E')}{(E-E')^2+[\Gamma(E')/2]^2}.$$

The broadening parameter Γ as a function of energy was expressed as $\Gamma(E)=\Gamma_0(1-E/E_F)^2$, $\Gamma_{e0}/\Gamma_{h0}=E_{Fe}/E_{Fh}$.[17,18] The level broadening is mainly due to Auger processes under the EHL Fermi surface. The parameter Γ_0 is a function of magnetic field, and can be varied when the formula is fitted to experimental data.

In calculating the EHL density of states, $D_{e,h}(E,B)$ in Eq. (2) was replaced with $D^*_{e,h}(E,B)$.

Values of g-factors of electrons and holes in the liquid phase in strained Ge were not measured. In a two-component Fermi-liquid, which applies to the EHL in strained germanium, they should be close to g-factors of free carriers. For a free electron, the g-factor is well known, and for the $[\sim 100]$ direction it is 1.57. The hole g-factor for Ge$[\sim 100]$, $g_h=-6.85\pm 0.06$, was taken from an investigation of Raman resonance in strained germanium.[19] The spin splittings for electrons and holes are $\Delta E_e=\mu_B g_e B[\mathrm{T}]=0.096$ meV and $\Delta E_h=\mu_B g_h B[\mathrm{T}]=0.348$ meV.

The lowest Landau levels, which determine the positions of the edges of conductance and valence bands, move with the magnetic field at the rates $(1/2B)(\hbar\omega_e-g_e\mu_B B)\approx 0.38$ meV/T (conductance band) and $(1/2B)\times(\hbar\omega_h+g_h\mu_B B)\approx 0.14$ meV/T (valence band).

Measured with respect to the band edges, the next Landau levels have energies $\hbar\omega_h[\mathrm{meV}]\approx 1.06B[\mathrm{T}]$ and $\hbar\omega_e[\mathrm{meV}]\approx 0.86B[\mathrm{T}]$. This means that for all fields $B>4$ T, only pairs of spin-split components of the lowest electron and hole Landau levels can lie below the respective Fermi levels.

The fit parameters were the EHL density n_{EHL} and Γ_0. The liquid temperature was equated to the excitonic gas temperature.

The density of states $D^*_{e,h}$ can be expressed as a sum of densities of states of spin-split levels, $D^*_{s_z,j_z}(E_{e,h},B)$:

$$D^*_{e,h}(E_{e,h},B)=\sum_{s_z,j_z} D^*_{s_z,j_z}(E_{e,h},B),$$

where s_z is the electron spin projection, and j_z is the respective hole spin component.

In approximating the line profile, one should take into account that the probabilities of recombination between electrons and holes from levels with different spin projections are different. Integral (2) can be written as the sum

$$I(h\nu)=\sum_{j,k}\alpha(s_z,j_z)I_{s_z,j_z}(h\nu),$$

$$I_{s_z,j_z}(h\nu)=A\int_0^\infty D^*_e(E_e,s_z)f(E_e,E_{Fe},T)\times D^*_h(E_h,j_z)f(E_h,E_{Fh},T)\,dE_e$$

with $s_z=\pm 1/2$ and $j_z=\pm 1/2$. The coefficients $\alpha(s_z,j_z)$ are dictated by the selection rules for the phonon-assisted dipole-allowed transition from point L_1 of the Brillouin zone to its center via intermediate state Γ_2. They are given by Asnin *et al.*[20] Using these parameters, we obtain $\alpha(1/2,1/2)=\alpha(-1/2,-1/2)=1/5$ and $\alpha(1/2,-1/2)=\alpha(-1/2,1/2)=4/5$.

Fits to the line profiles using this procedure for two magnetic fields of $B=5.5$ and 13.5 T are shown in Fig. 8a,b. The arrows mark the chemical potential, renormalized gap width $E_g'^{\mathrm{LA}}$, and the edge of the exciton line obtained by approximating its profile with the expression $\sqrt{E}\,\exp(-E/k_BT)$, with due account of the monochromator instrumental function.

It is possible to obtain a satisfactory fit to the experimental EHL line profile by assuming that there is one Landau level for electrons and one for holes below the Fermi levels. The parameters E_{Fe} and E_{Fh} increase by 50% as the magnetic field grows from 5.5 to 13.5 T (from 1.1 to 1.6 meV and from 0.11 to 0.17 meV, respectively) (Fig. 9a). The observed increase in the EHL line width is due to both a change in $E_{Fe,h}$ and an increase in $\Gamma_0(B)$ (Fig. 9a,b). It is notable that the resulting Γ_0 is large for a one-particle model. Whereas for $B=4$ T we have $\Gamma_{0h}<E_{Fh}$, for $B=13.5$ T the ratio $\Gamma_{0h}/E_{Fh}\sim 2$. Such high values of Γ_0 were needed for an accurate description of the low-energy tail of the EHL line. Nonetheless, our analysis of the line profile yields, in our opinion, correct values of n_{EHL} and ϕ to within the speci-

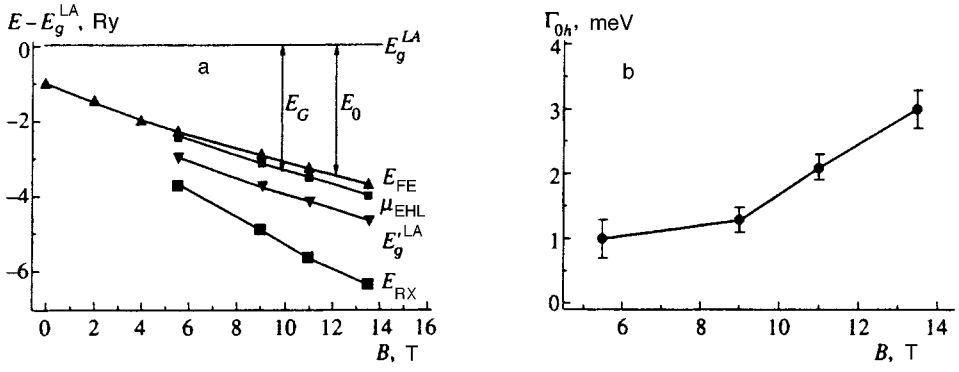

FIG. 9. a) EHL parameters as functions of magnetic field. The energy is measured with respect to E_g^{LA}, i.e., the gap width minus the LA-phonon energy, E_{FE} is the low-energy edge of the FE line, $E_g'^{\mathrm{LA}}$ is the renormalized gap width in the liquid, E_{RX} is the energy of the red edge of the EHL line (see caption of Fig. 6), E_0 is the exciton binding energy, E_{G} is the EHL ground state energy, μ_{EHL} is the liquid chemical potential. The energy is measured in units of Ry=2.8 meV in Ge[$\sim$100]; b) function $\Gamma_{0h}(B)$ derived from approximations to the EHL line profile.

fied uncertainties, because they depend weakly on Γ_0 (as Γ_{0h} varies from 0.1 to 3 meV, n_{EHL} decreases by less than 10%, and the binding energy is dictated by the high-energy shoulder of the line, which is essentially independent of Γ_0).

The density n_{EHL} versus B derived from this approximation is given in Fig. 10a. The EHL density approximately triples in the field range $B=5.5-14$ T (from $1.25\cdot10^{16}$ cm^{-3} to $3.6\cdot10^{16}$ cm^{-3}). This almost linear increase in the density in Ge[$\sim$100] (Fig. 10a) with a small increase in $E_{Fe,h}$ is largely due to the increase in the capacity of the Landau level, which is proportional to B.

The energy E_{G} of the liquid ground state per electron–hole pair as a function of B is plotted in Fig. 9a.

An increase in the ground state binding energy and EHL density with magnetic field was predicted by Keldysh and Onishchenko.[4] Their theoretical analysis was based on the assumption that the field strength is so high that all electrons and holes are in the lowest Landau levels. The case of strained Ge[$\sim$100] approximately satisfies the conditions of Keldysh's theory. Our results are in good agreement with calculations based on this model.

From the approximation of the line profile, we have derived the binding energy ϕ as a function of B. The value of ϕ was given by the separation between the liquid chemical potential and the low-energy edge of the FE line (Fig. 8). The function $\phi(B)$ is plotted in Fig. 10b. The binding energy ϕ increases from 0.4 ± 0.1 to 0.8 ± 0.1 meV in the field range 5.5–14 T.

Since the ground state energy of the metallic liquid grows as a power of the magnetic field in the limit $\gamma\to\infty$, while the exciton binding energy grows logarithmically, the condensed phase is more stable than the excitonic gas in a strong magnetic field. In intermediate fields, this is not so obvious (see, for example, Silin's paper in Ref. 11). In unstrained Ge in fields $B<19$ T (the ratio $\gamma<1$, and there are several Landau levels below the Fermi level), Störmer and Martin[16] did not detect an increase in ϕ. Given this fact, the emergence of the EHL line in the luminescence spectrum of strained germanium and its rapid growth with magnetic field is, in our opinion, all the more interesting. It seems that, given the small absolute value of the binding energy and the proximity to the phase boundary, even a small change in ϕ leads to fast growth in the fraction of the liquid phase (Figs. 4 and 5a). Our results indicate that the growth in ϕ becomes appreciable at $\gamma\sim1$.

The presence of a fairly wide biexciton line in strongly strained Ge[$\sim$100] at $B<4$ T did not allow us to trace the evolution of the EHL line in low magnetic fields and address the question of whether the EHL exists at low temperatures in zero magnetic field, as opposed to whether it emerges at a critical value B_c. According to theoretical estimates,[11] EHL should exist in single-valley Ge(1:1), i.e., when both electrons and holes occupy only one spin-degenerate level, even in the limit of infinite pressure. Its density at $T=0$ K is $n_{\mathrm{EHL}}\approx6\cdot10^{15}$ cm^{-3}, and the critical temperature $T_c=2.2-2.5$ K. In the case of a finite strain, T_c will be higher. In our analysis of the line profile, we assumed phase equilibrium between the gaseous and condensed phases. The EHL temperature was assumed equal to the exciton gas temperature. In reality, temperature of an electron–hole drop is higher than that of the ambient excitonic gas,[15] and can be higher than the temperature at which the existence of stable electron–hole drops is possible.

At the same time, the critical temperature of the gas–

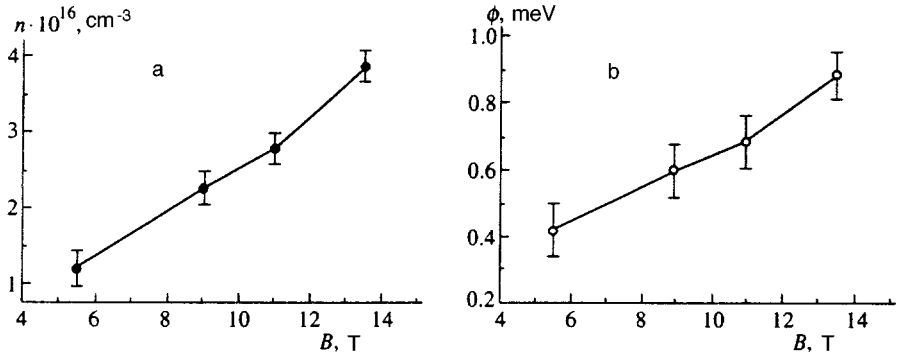

FIG. 10. a) Density $n_{\mathrm{EHL}}(B)$ and b) EHL binding energy $\phi(B)$ in Ge[$\sim$100] at $P=2.8$ kbar and $T=1.9$ K derived from approximations to the luminescence line.

liquid transition increases with both ϕ and n_{EHL} (in the strong-field limit $T_c \propto B^{2/7}$; see Ref. 21). Therefore an increase in T_c will shift the phase equilibrium boundary and lead to stabilization of droplets. Then the EHL line should emerge abruptly in the luminescence spectrum. Emergence of EHL in magnetic fields higher than 2 T was detected in InSb,[22] where in the absence of a magnetic field, no EHL exists at liquid-helium temperatures. In this case, electrons and holes also occupy only the lowest Landau levels. The condensed phase density n_{EHL} was measured[22] to be about 10^{16} cm^{-3}, and both the binding energy and density of the liquid increased with B. Note also that the line-fit parameter Γ responsible for broadening of electron and hole levels in a magnetic field was comparable to the Fermi energy of the latter.

It seems more feasible that, in strained Ge near the phase boundary, there is initially a small quantity of stable EHL droplets, which grow rapidly when the magnetic field is turned on. In low magnetic fields below 4 T, the EHL binding energy $\phi \approx 0.3$ meV, i.e., it is in fact equal to the excitonic molecule binding energy $\phi_{\mathrm{EM}} = 0.27 \pm 0.06$ meV.[14] In the absence of a magnetic field, the binding of excitons into biexcitons competes with exciton condensation into an electron–hole liquid. A magnetic field destroys biexcitons, thus leading to an increase in the liquid fraction.

Assuming that in the case of Ge[~ 100] ($P = 2.8$ kbar) under discussion, the width of the liquid line at $B = 0$ equals its value at $B = 4$ T, let us estimate n_{EHL} at zero magnetic field. Given that the EHL line width is the sum of the Fermi energies $E_{Fe} + E_{Fh}$, we have $n_{\mathrm{EHL}} \simeq 10^{16}$ cm^{-3}. This result is in agreement with n_{EHL} derived from approximations of the luminescence line profile (Fig. 9a). An electron–hole liquid of such low density is quite feasible. Kukushkin and Kulakovskii[23] investigated EHL with $n_{\mathrm{EHL}} \simeq 0.8 \cdot 10^{16}$ cm^{-3} ($r_s \approx 1.65$) in uniaxially strained Ge[~ 100] at zero magnetic field, whereas in unstrained Ge at $T = 2$ K the EHL density $n_{\mathrm{EHL}} \simeq 2.4 \cdot 10^{17}$ cm^{-3} and $r_s \approx 0.5$.

Strain nonuniformity, which is always present in experiments, should also stabilize EHL drops in strained germanium in zero magnetic field.

The authors are indebted to V. D. Kulakovskii for helpful discussions and interesting remarks. This work was supported by the Russian Fund for Fundamental Research, project 29-02-17535.

[1] R. J. Elliot and R. Loudon, J. Phys. (France) **15**, 196 (1960).
[2] G. V. Godiak, Yu. E. Lozovik, and M. S. Obrekht, Fiz. Tverd. Tela (Leningrad) **25**, 1063 (1983) [Sov. Phys. Solid State **25**, 613 (1983)].
[3] Y. Kuramoto and M. Morimoto, J. Phys. Soc. Jpn. **44**, 1759 (1978).
[4] L. V. Keldysh and T. A. Onishchenko, JETP Lett. **24**, 59 (1976).
[5] S. T. Chui, Phys. Rev. B **9**, 3438 (1974).
[6] V. B. Timofeev and A. V. Chernenko, JETP Lett. **61**, 617 (1995).
[7] A. V. Chernenko, Prib. Tekh. Eksp. **5**, 141 (1996).
[8] G. L. Bir and G. E. Pikus, *Symmetry and Deformation Effects in Semiconductors*, [in Russian], Nauka, Moscow (1972).
[9] I. Balslev, Phys. Rev. **143**, 636 (1966).
[10] J. V. Hensel and K. Suzuki, Phys. Rev. B **9**, 4219 (1974).
[11] *Electron-Hole Drops in Semiconductors*, L. V. Keldysh and C. Jeffries (eds.), Nauka, Moscow (1988).
[12] T. G. Tratas and V. M. Édel'shteĭn, Zh. Éksp. Teor. Fiz. **81**, 696 (1981) [Sov. Phys. JETP **54**, 372 (1981)].
[13] I. V. Kukushkin and V. D. Kulakovskiĭ, JETP Lett. **32**, 304 (1980).
[14] V. D. Kulakovskiĭ and I. V. Kukushkin, Zh. Éksp. Teor. Fiz. **81**, 684 (1981) [Sov. Phys. JETP **54**, 366 (1981)].
[15] T. Rice, J. Hensel, T. Philips, and G. A. Thomas, *Electron–Hole Liquid in Semiconductors*, Academic Press, New York (1977).
[16] H. L. Störmer and R. W. Martin, Phys. Rev. B **20**, 4213 (1979).
[17] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics*, [in Russian], Fizmatgiz, Moscow (1962).
[18] V. D. Kulakovskiĭ, V. B. Timofeev, and V. M. Édel'shteĭn, Zh. Éksp. Teor. Fiz. **74**, 372 (1978) [Sov. Phys. JETP **47**, 193 (1978)].
[19] J. S. Hensel and K. Suzuki, Phys. Rev. Lett. **40**, 838 (1969).
[20] V. M. Asnin, G. L. Bir, Yu. N. Lomasov, G. E. Pikus, and A. A. Rogachev, Zh. Eksp. Teor. Fiz. **71**, 1600 (1976) [Sov. Phys. JETP **44**, 838 (1976)].
[21] T. A. Onishchenko, Dokl. Akad. Nauk. SSSR **235**, 78 (1977) [Sov. Phys. Dokl. **22**, 383 (1977)].
[22] I. V. Kavetskaya, Ya. Ya. Kost', N. N. Sibel'din, and V. A. Tsvetkov, JETP Lett. **36**, 311 (1982); I. V. Kavetskaya, N. N. Sibeldin, and V. A. Tsvetkov, Solid State Commun. **97**, 157 (1996).
[23] I. V. Kukushkin and V. D. Kulakovskiĭ, Zh. Éksp. Teor. Fiz. **82**, 900 (1982) [Sov. Phys. JETP **55**, 528 (1982)].

Translation was provided by the Russian Editorial office.

Direct and spatially indirect excitons in GaAs/AlGaAs superlattices in strong magnetic fields

A. I. Tartakovskiĭ and V. B. Timofeev

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

V. G. Lysenko

Institute of Microelectronics and High-Purity Materials, 142432 Chernogolovka, Russia

D. Birkedal and J. Hvam

Microelectronics Center, The Technical University of Denmark, DK 2800 Lyngby, Denmark
(Submitted 6 February 1997)
Zh. Éksp. Teor. Fiz. **112**, 1106–1118 (September 1997)

Luminescence and luminescence excitation spectra are used to study the energy spectrum and binding energies of direct and spatially indirect excitons in GaAs/AlGaAs superlattices, with different widths of the electron and hole minibands, located in a high magnetic field perpendicular to the heterolayers. It is found that the ground state of the indirect excitons formed by electrons and holes and spatially separated between neighboring quantum wells lies between the 1*s* ground state of the direct excitons and the continuum threshold for dissociated exciton states in the minibands. Indirect excitons in superlattices have a significant oscillator strength when the binding energy of the exciton exceeds the order of the width of the resulting miniband. The behavior of the binding energy of direct and indirect heavy hole excitons during changes in the tunneling coupling between the quantum wells is established. It is shown that a strong magnetic field, which intensifies the Coulomb interaction between the electron and hole in an exciton, weakens the bond in a system of symmetrically bound quantum wells. The spatially indirect excitons studied here are analogous to first order Wannier–Stark localized excitons in superlattices with inclined bands (when an electrical bias is applied), but in the present case the localization is of purely Coulomb origin. © *1997 American Institute of Physics.* [S1063-7761(97)02209-9]

1. INTRODUCTION

The properties of excitons have continued to be at the center of scientific interest in studies of optical and electronic phenomena in artificially prepared superlattices based on semiconducting heterostructures.[1–13] From the very beginning, interest in these quantum mechanical structures has been associated with research on the gradual transformation of their properties from the quasi-two-dimensional limit (single quantum well) to three-dimensional solids (superlattice with strong tunnelling coupling between quantum wells, or, in other words, fairly wide electron and hole minibands). One of the central questions in this area has concerned the structure of the exciton states corresponding to the lower (absolute minimum M_0) and upper (saddle M_1) points with zero slope in the miniband spectrum. It is appropriate to mention here, however, that there is still some controversy over whether ''saddle'' excitons have actually been observed at the upper boundary of the miniband spectrum in superlattices.[8,9]

When a sufficiently high electric field is applied perpendicular to the superlattice layers, single-particle electron and hole states become localized in individual quantum wells. This sort of Wannier–Stark electron localization was identified and studied some time ago,[10,11] and this has, in turn, stimulated experiments to observe Bloch oscillations[12] and the electromagnetic radiation produced by these oscillations.[13] All of these oscillatory phenomena are closely related to the spatially indirect nature of Vannier-Stark localized states, including exciton states.

Spatially indirect excitons can, however, exist outside bound quantum mechanical systems with broken inversion symmetry. Theoretical analyses[4,5,14] of the complete exciton Hamiltonian for symmetrically bound superlattices have yielded stable solutions for the intrinsic energy spectrum of spatially indirect excitons, i.e., when the electron and hole of an exciton are separated by a barrier. Recall that the complete version of the exciton Hamiltonian in this case does not have an exact solution, since the spatial coordinates in the planes of the layers and in the axial direction are not separable. Hence, in principle, the problem must be solved approximately. Thus, superlattice exciton states have been sought[4] using a reduced basis of wave functions corresponding to a single electron–hole miniband. This is a reasonably good approximation when the width of the miniband is of the order of the exciton binding energy. The principal result of Refs. 4 and 14 is then that the Coulomb interaction can localize an electron and hole in neighboring symmetrically bound quantum wells, and this leads ultimately to the formation of a spatially indirect exciton.

The formation of indirect excitons in superlattices is very close in nature to the analogous events which take place in symmetric double quantum wells.[15–20] As in double quantum wells, the bound exciton states in superlattices (direct *D*

and spatially indirect I) are significantly modified, i.e., their binding energy changes and the oscillator strengths among these states are redistributed dramatically when the relationship among the exciton binding energies $E_{ex}(D)$ and $E_{ex}(I)$ and the width of the electron and hole minibands $\Delta_e+\Delta_h$ changes. Following the close analogy with symmetric double quantum wells,[20] the exciton states in superlattices can be analyzed by introducing the phenomenological parameter $\alpha=(\Delta_e+\Delta_h)/[E_{ex}(D)-E_{ex}(I)]$.

Thus, the newly introduced parameter α characterizes the tunnelling bond between the quantum wells relative to the Coulomb interaction in the exciton. In superlattices with wide barriers (narrow minibands), $\alpha \ll 1$. In this limit the exciton states are, by their nature, predominantly direct D and indirect I. In this case indirect excitons are extremely weakly bound and their oscillator strength is much lower than that of a direct exciton. For reasons of symmetry (the presence of a center of inversion in the system), all exciton states (direct and spatially indirect) are classified as symmetric or antisymmetric. In order of increasing energy, these states are ordered as follows: D-symmetric, D-antisymmetric, I-antisymmetric, and I-symmetric. Only symmetric exciton states are optically active (observed in spectra). It is perfectly evident that the splitting in a spectrum between D- and I-exciton states is determined by the difference in the binding energies of these excitons. In superlattices with narrow barriers, $\alpha \gg 1$, the exciton states have a mixed direct–indirect character. In this limit, where the Coulomb interaction is negligibly small, the exciton states are classified as the product of single-particle electron–hole states of the corresponding symmetry.

The case of an intermediate bond, where the width of the miniband is of the order of the Coulomb binding energy in the exciton, is also of interest. In this case the oscillator strength of an indirect exciton can be as much as a few percent of the corresponding value for a direct exciton.[4–14]

Finally, in the spectra of superlattices, indirect excitons must lie between the ground state of a direct exciton and its continuum of dissociated states (the minimum point M_0).

A magnetic field B perpendicular to the superlattice layers has a significant effect on exciton states. In the limit of sufficiently high magnetic fields, the binding energies for direct $E_{ex}(D)$ and indirect $E_{ex}(I)$ excitons can be approximated by the respective expressions

$$E_{ex}(D) \propto e^2/\epsilon l_0,$$

$$E_{ex}(I) \propto e^2/\epsilon (l_0^2+d^2)^{1/2},$$

where $l_0=(\hbar c/eB)^{1/2}$ is the magnetic length and d is the effective spatial distance between an electron and hole in the axial direction (or the superlattice constant). Evidently, the binding energy for a direct exciton $E_{ex}(D) \propto B^{1/2}$, while $E_{ex}(I)$ increases more slowly under these conditions and approaches a constant value in the limit of strong fields. Thus, a magnetic field increases the splitting between the ground state of the direct and spatially indirect excitons and, thereby, for fixed tunnel coupling in superlattices, shifts the symmetry of the bound system of quantum wells toward a weak bond. Here it is assumed that in the absence of special resonance conditions, a vertical magnetic field has no effect on the width of the minibands, to a first approximation.

In this paper we have set ourselves the task of demonstrating that indirect excitons are formed in symmetrically bound quantum systems (superlattices) and that it is not necessary to break the inversion symmetry in order to observe them. For this purpose we have studied the energy spectrum of direct and spatially indirect excitons in superlattices as the tunnel coupling (barrier width) between quantum wells is varied. The use of a sufficiently strong magnetic field, which enhances the Coulomb bond in excitons, has made it possible to affect the character of the bond in superlattices.

The article is organized as follows. In Sec. 2, the structures and techniques used in the experiment are described. Section 3 deals with the behavior of the bond energy and oscillator strengths of direct and spatially indirect excitons as the barrier widths between quantum wells in superlattices are varied. The magnetooptics of direct and spatially indirect excitons is introduced in Sec. 4. The unexpected, to us, features of the behavior of excited exciton states in a magnetic field are described and a qualitative interpretation of these features is given in Sec. 5. Section 6 concludes the article.

2. STRUCTURES AND EXPERIMENTAL TECHNIQUES

In this work we used superlattices based on $GaAs/Al_xGa_{1-x}As$ ($x=0.3$) heterostructures grown by molecular beam epitaxy on an insulating GaAs substrate oriented along the [001] crystallographic direction. The structures were not specially doped, so a ''flat band'' regime was realized in all the superlattices. Five superlattices were studied; these had the same GaAs quantum well width $L_w=80$ Å and different AlGaAs barrier widths L_b between the wells—specifically, $L_b=20$, 30, 50, 100, and 200 Å. When the barrier width was varied within these limits, the width of the resulting electron and hole minibands, $\Delta_e+\Delta_h$, ranged from 25 to 0.1 meV.

Each structure had 20 periods $d=L_w+L_b$. Accurate parameters of the superlattices were determined using x-ray diffraction. Optimized epitaxial growth conditions provided superlattices of high quality, with good reproducibility of the layer thicknesses and perfection of the interfaces between layers. The high quality of the test structures is attested to by the exciton line widths in the luminescence and luminescence excitation spectra (about 1 meV), as well as by the magnitude of the Stokes shift of the lines in the luminescence and absorption or reflection spectra (less than 0.5 meV).

The samples were mounted freely in a helium cryostat inside a superconducting solenoid. The maximum attainable field was 14 T. In all the experiments, the magnetic field was perpendicular to the superlattice layers. The luminescence and luminescence excitation spectra were measured using a wavelength tunable titanium sapphire laser pumped by all the emission lines of a cw argon laser. Fiber optics were used to excite and collect the luminescence signal from the sample. All the spectra were analyzed using a double monochromator which provided adequate spectral resolution (better than 0.1 meV). In the luminescence excitation spectral measurements the output slit of the monochromator was

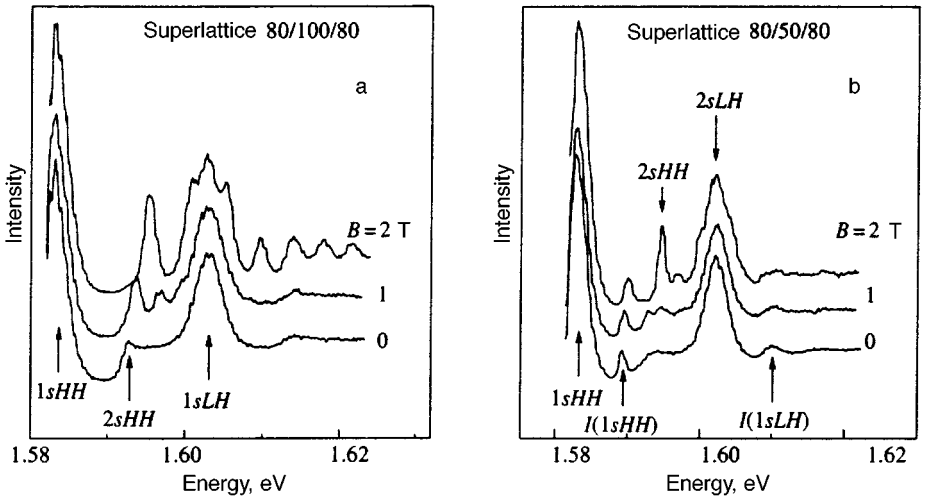

FIG. 1. Luminescence excitation spectra of the 80/100/80 (a) and 80/50/80 (b) superlattices taken in magnetic fields of up to 2 T.

tuned to a narrow segment of the spectrum on the long wavelength wing of the luminescence line of the lowest heavy hole exciton state ($1sHH$) (at the maximum of the luminescence line or about 1 meV from it). The locations of the peaks in the discrete luminescence excitation spectra were independent of the spectral detection region to within the width of the luminescence line of the exciton ground state. At the monochromator output, the luminescence signal was detected by a photomultiplier with a cooled GaAs photocathode, and then recorded in photon counting mode.

3. DIRECT AND INDIRECT EXCITONS IN SUPERLATTICES IN THE ABSENCE OF A MAGNETIC FIELD

a. Above all, we consider the behavior of the bond energy of direct and indirect heavy hole excitons as the width of the quantum barrier in superlattices is varied in the absence of a magnetic field. This question can be addressed by analyzing the luminescence excitation spectra. To illustrate the procedure for determining the binding energy of the heavy hole exciton ground state, we turn to Fig. 1, which shows luminescence excitation spectra for two superlattices, with barrier widths $L_w = 100$ Å and 50 Å (hereafter these are referred to as superlattices 80/100/80 and 80/50/80, respectively). The strongest lines in the spectra are those of the ground state of excitons with heavy ($1sHH$) and light ($1sLH$) holes (denoted by arrows in the figure). The dissociation region of the HH-excitons shows up as a distinct continuum with a sharp ''red'' threshold above which the $1sLH$-exciton stands out. The first excited state of the heavy hole exciton ($2sHH$) has the form of a feature (weak peak) on the ''red'' boundary of this continuum. This feature grows into a distinct and fairly intense line even in low magnetic fields of 1–2 T. In all the structures that were studied, it is possible to measure an energy difference corresponding to the spectral positions of the $2sHH$- and $1sHH$-excitons, $\Delta(2s-1s)$. Note that the $2sHH$-exciton state is very weakly bound in the cases of interest to us, and lies within 1 meV below the continuum threshold, which is close to the experimental accuracy with which the spectral position of the lines (which are 1–1.5 meV wide), is determined. Thus, the experimentally measured value of $\Delta(2s-1s)$ is within 15% of the binding energy of a heavy hole exciton. The corresponding measurements of $\Delta(2s-1s)$ as a function of the barrier size L_w are shown in Fig. 2. In this figure the dashed curve shows the variation in the combined width of the electron Δ_e and hole Δ_h minibands with changing barrier width L_w, calculated in the one-electron approximation using the Kronig–Penney model.[21] When the barrier width is varied over these limits, the exciton structure changes from quasi-two-dimensional ($L_w = 200$ Å) to a situation that is close to three dimensional ($L_w = 20$ Å). Accordingly, the binding energy of the heavy hole exciton decreases, in agreement with an earlier paper.[2] Qualitatively the same result is obtained for light hole excitons, although the measurement accuracy in this case is much lower.

b. In superlattices, starting at barriers having $L_b = 50$ Å and less, a new line, $I(1sHH)$, that is not associated with the spectrum of the direct excitons is seen clearly in the luminescence excitation spectra between the $1sHH$- and $2sHH$-lines (see Fig. 1(b)). We previously briefly reported the first observation of this line in an 80/50/80 superlattice.[22]

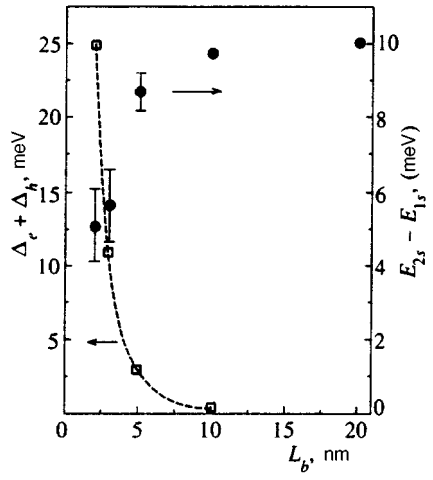

FIG. 2. The difference in the energies of the first excited and ground states of an exciton, $E_{2s}-E_{1s}$ (right scale, black circles) and the magnitude of the total miniband (left scale, dashed curve) as functions of the width of the barrier between wells.

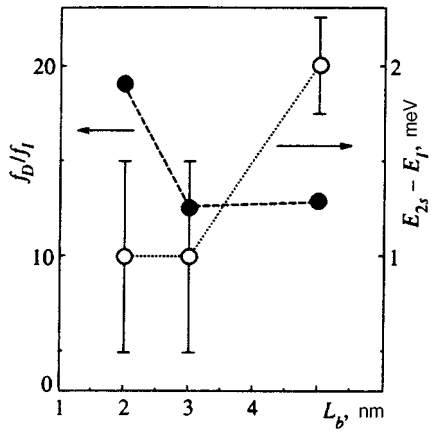

FIG. 3. The energy difference $E_{2s}-E_I$ (right scale, white circles) and relative oscillator strength of an indirect exciton (left scale, black circles) as functions of the width of the barrier between wells.

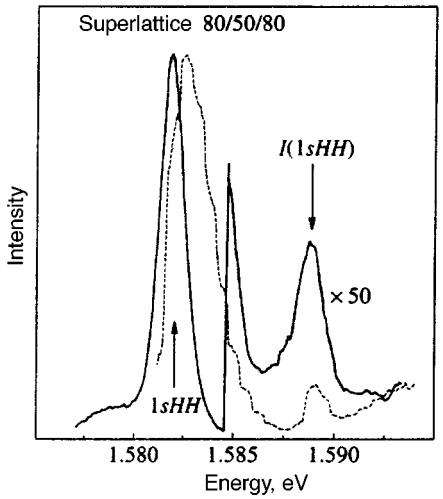

FIG. 4. Luminescence (smooth curve) and luminescence excitation (dashed curve) spectra of the 80/50/80 superlattice without a magnetic field.

In the 80/50/80 superlattice this line corresponds to the ground state of a heavy hole exciton with a spatially separated electron and hole at a distance corresponding to a single period. Thus, in this case we are dealing with a spatially indirect exciton that is localized in the axial direction within the confines of a single superlattice period. Unlike Wannier–Stark localization, in this case the localization is exclusively of Coulomb origin. The indirect exciton state is bound more weakly than the corresponding state of the direct exciton. The oscillator strength of the I-exciton is highest in superlattices with intermediate bonds. Thus, in the 80/30/80 superlattice, where α is roughly equal to unity, the oscillator strength of an indirect exciton is about 10% of that of a direct heavy hole exciton. As the barrier width is reduced ($L_b=20$ Å and below), the wave function of the indirect exciton becomes more extended in the z direction and may even encompass several superlattice periods. In this limit the indirect exciton states become more and more delocalized, and ultimately merge with the dissociation continuum of the direct excitons.[4] Figure 3 illustrates the behavior of the binding energy of the indirect exciton, as well as its oscillator strength measured relative to the direct exciton, as the barrier width is changed. It is clear from this figure that for superlattices with quantum well widths $L_w=80$ Å, the optimum conditions for observing spatially indirect excitons occur at quantum mechanical barrier widths of about 30 Å.

In concluding this section, we note that qualitatively similar behavior of the binding energies and oscillator strengths is manifested by spatially indirect light hole excitons (the $I(1sLH)$ line in the figures), but the optimum conditions for observing them occur at relatively large barrier widths ($L_b=50$ Å) because of the larger width of the hole miniband in the case of light holes.

c. Bound states corresponding to spatially indirect excitons are also observed in the luminescence spectra. For the 80/50/80 superlattice this is illustrated by Fig. 4, where it can be seen that the $I(1sHH)$ lines in the luminescence and luminescence excitation spectra coincide. By comparing data on the luminescence spectra for the $1sHH$ and $I(1sHH)$ excitons, one can easily conclude that the indirect excitons are not in thermal equilibrium with the direct excitons. Note that under these same conditions, we do not see any appreciable sign of the excited $2sHH$ state of the direct exciton in the luminescence spectra. This indicates that direct excitons undergo thermal relaxation much more rapidly. In fact, the population of the states corresponding to the indirect excitons is many orders of magnitude higher than the thermal equilibrium concentration. This means that relaxation of the indirect I-exciton to the lowest $1sHH$-exciton state in this system is a relatively slow process. This conclusion is not surprising, given that indirect excitons can relax into the lowest state, which corresponds to direct excitons, by nonresonant tunnelling with phonon emission, which really is a slow process for these widths of the barrier between the quantum wells.[23]

4. EXCITATION SPECTRA OF DIRECT AND SPATIALLY INDIRECT MAGNETOEXCITONS

The nature of the indirect excitons observed in the luminescence and luminescence excitation spectra is confirmed by the behavior of their spectra in a magnetic field. In the limit of a weak magnetic field, when the cyclotron energy is much lower than the Rydberg constant for the excitons, the diamagnetic shift of an exciton is proportional to the cross sectional area of the exciton perpendicular to the magnetic field B. For two-dimensional excitons the Bohr radii $a_n=a_0(2n+1)$ for $n>1$, where n is the principal quantum number. The magnitude of the diamagnetic shift itself is proportional to $\delta E_{\rm dia}\propto a_0^2(2n+1)^2\propto 1/E_{ex}$. Thus, for two-dimensional $1s$ and $2s$ excitons, the ratio of the diamagnetic shifts is 9. The indirect excitons observed in the superlattices with barrier widths $L_b=20$, 30, and 50 Å are more weakly bound than the direct $1s$-exciton, but more strongly than the $2s$-exciton state in the same structures. Thus, the diamagnetic shift for an indirect exciton should be substantially smaller than for direct $2s$-excitons. On the other hand, a sufficiently strong magnetic field should transform a symmetric tunnelling-coupled system of quantum wells toward weaker bonds. This should be accompanied by a reduction in

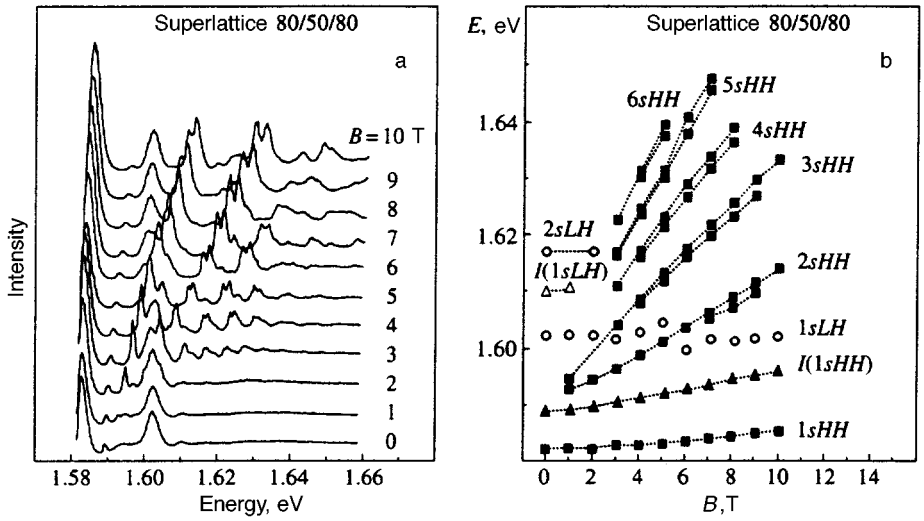

FIG. 5. Luminescence excitation spectra of the 80/50/80 superlattice for various magnetic fields (a), and the energies of the peaks in these spectra as a function of magnetic field (b).

the oscillator strength of the direct exciton, owing to the weakening mixing of the direct exciton states.

Figures 5(a), 6, and 7 show luminescence excitation spectra for the superlattices with barrier widths $L_b = 50$, 30, and 20 Å, measured in magnetic fields of up to 14 T with a step size of 1 T. The strongest lines in the spectra are those of the direct $1sHH$ and $1sLH$ excitons. As the magnetic field is increased, these exciton states exhibit a diamagnetic shift which is greater in magnitude for excitons with a heavy hole, since this state is more weakly bound than the $1sLH$ exciton. As the barrier width is reduced ($L_b = 30$, 20 Å), the diamagnetic shift of the $1sLH$ exciton is essentially zero, because the effect of the superlattice is greatest for exciton states with a light hole.

Besides the direct exciton ground states, which transform into the magnetoexciton states with the lowest Landau level ($N = 0$) as the magnetic field is increased, in the luminescence excitation spectra we have observed magnetoexcitons corresponding to higher Landau states ($N > 1$). Only those magnetoexcitons which consist of bound states of an electron and hole belonging to one and the same diamagnetic quantization level are optically active in the spectra. The corresponding exciton energies for the direct and spatially indirect excitons are

$$E^N_{D(I)} = E_g + \hbar(\omega^e_c + \omega^h_c)\left(N + \frac{1}{2}\right) - E_{ex}(D,I),$$

where $\hbar\omega^e_c$ and $\hbar\omega^h_c$ are the cyclotron energies, $E_{ex}(D,I)$ are the binding energies of the direct (bent) excitons corresponding to Landau level N, and E_g is the distance between the electron and hole size-quantization levels. Magnetoexciton states up to quantum numbers $N = 6$ were reliably identified in the luminescence excitation spectra of the superlattices studied here.

The $2sHH$ exciton is observed as a well-defined line, even in weak magnetic fields (about 2 T). With increasing

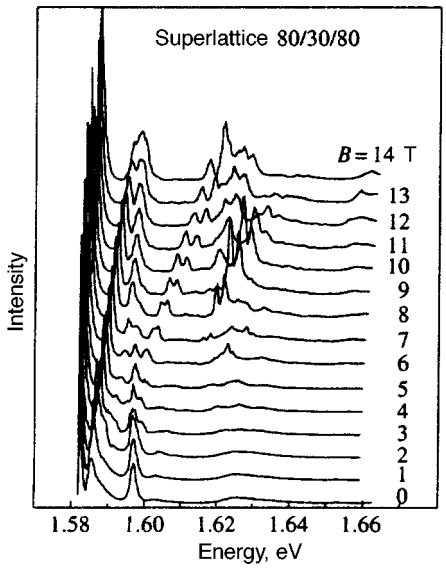

FIG. 6. Luminescence excitation spectra of the 80/30/80 superlattice for various magnetic fields.

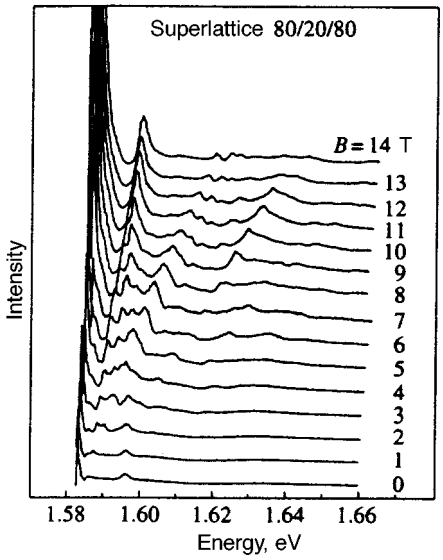

FIG. 7. Luminescence excitation spectra of the 80/20/80 superlattice for various magnetic fields.

field, the $2sHH$ exciton, because of its large diamagnetic shift, overlaps the spectral position of the $1sLH$ light hole exciton. In the overlap region of these exciton states in all the superlattices we have studied, one can see a dramatic redistribution of the oscillator strengths of the $2sHH$- and $1sLH$-lines. The observed anticrossing effect is the result of fairly strong mixing of light and heavy hole magnetoexciton states.

We now consider the diamagnetic properties of the direct and indirect excitons. Approximating the spectral shift of the exciton line as a function of the magnetic field at low fields by a quadratic dependence, we find that in the 80/50/80 superlattice the magnitude of the diamagnetic shift is 37 μeV/T^2 for the $1sHH$ exciton state. This value is in good agreement with data for a 75-Å-wide quantum well in Ref. 2. As the barrier width in the superlattices (and therefore the direct exciton binding energy) are reduced, the magnitude of the diamagnetic shift of the $1sHH$ state increases. Thus, in the 80/20/80 superlattice the diamagnetic shift of the $1sHH$ exciton is roughly 64 μeV/T^2. Furthermore, the diamagnetic shift of the $2sHH$ exciton is almost an order of magnitude greater than for the ground state in all the superlattices that were studied, in close accord with the estimates given above.

At the same time, the diamagnetic corrections to the ground state energy of a spatially indirect exciton are substantially greater than for a $1sHH$ exciton, and much smaller than for a $2sHH$ exciton. This correction is approximately 100 μeV/T^2 in the 80/50/80 superlattice. Thus, it is perfectly obvious that the $I(1sHH)$ exciton has no connection with excited states of the direct excitons. We conclude that the $I(1sHH)$ line corresponds to the $1s$ ground state of an indirect exciton with a spatially separated electron and hole within the confines of the superlattice period.

Because of the high barrier, an indirect exciton in the 80/100/80 superlattice is very weakly bound and its oscillator strength is very low. Thus, this exciton state lies up against the dissociation region of the heavy hole excitons, and is not allowed in the luminescence excitation spectra at low magnetic fields. However, ''traces'' of the indirect magnetoexciton are observed at high magnetic fields, $B>11$ T, in a region that essentially coincides with the location of the Landau level $N=0$ with light holes.

In superlattices with narrow barriers, $L_b=30$, 20 Å, a satellite fine structure is observed in the region of the principal indirect exciton line (see Figs. 6 and 7). As the magnetic field is increased, this structure merges with the dissociation region of the heavy hole direct excitons, so that only one line of the indirect exciton remains in the spectrum. We attribute the emergence of this structure and its behavior in a magnetic field in superlattices with narrow barriers to spatially indirect excitons in which the electron and hole are separated by two or three superlattice periods, in accord with earlier theoretical predictions.[4]

Spatially indirect excitons in which light holes are also bound are observed in the luminescence excitation spectra of the 80/50/80 superlattice. The corresponding line is labeled $I(1sLH)$ in Fig. 1(b). The state corresponding to this line exhibits a substantially larger diamagnetic shift than for the ground state of the direct exciton with a light hole, $1sLH$, but considerably smaller than for the excited $2sLH$ state. As the magnetic field is reduced ($B\to 0$), the spectral position of the $I(1sLH)$ line can be extrapolated to the threshold of the dissociation continuum for the light hole exciton states. All these arguments convince us that $I(1sLH)$ is associated with spatially indirect light hole excitons.

5. SPECTRAL FEATURES OF EXCITED MAGNETOEXCITON STATES

We have observed several surprising features in the magnetic field variation of the excited state spectra of magnetoexcitons. Here we discuss two of the most striking of them.

a. In the luminescence excitation spectra of all the superlattices we have studied, we found that at sufficiently high magnetic fields, the heavy hole magnetoexciton states corresponding to Landau levels $N\geqslant 1$ have a distinct Zeeman splitting doublet structure (see, e.g., Fig. 6). Special experiments in a Faraday geometry with analysis of the circular polarization of the doublet components showed that the lower-energy component corresponds to σ^- polarization and that on the higher, to σ^+. The observed splitting is associated with quantization of the projections of the electron and hole spins in the exciton, since in this case we are dealing with magnetoexciton s states.

The magnitude of the doublet splitting itself was surprisingly large, if we adopt published values for the electron and hole g-factors.[24] The reason for the anomalously large Zeeman splitting may lie in the fact that in systems with confinement, other states with suitable symmetry, but with much larger orbital angular momentum (e.g., d, f, and other states), may mix with the s states of an exciton.[25] Because of strong mixing with a state that has large angular momentum and because of spin–orbit interaction, the effective g-factor of the hole in the s exciton state increases.[26] This assumption is true if when confinement is weakened (i.e., when the barrier width is reduced in superlattices), the magnitude of the doublet splitting decreases.

Conservation of angular momentum implies that the components of the effective spin of the hole m_J and electron s in the exciton are related to the angular momentum of the photon m_{ph} by $m_{ph}=s+m_J$. For the $\sigma^{\pm}$ optical polarizations, $m_{ph}=\pm 1$. For optical transitions it is, of course, necessary to consider conservation of the parity of the states in the direction of size quantization. With these remarks, the observed doublet structure in the Zeeman splitting of the excited heavy hole s magnetoexciton states can be explained by the optical transition scheme represented in Fig. 8. We further assume that the Zeeman Hamiltonian of an exciton in a magnetic field oriented along the z axis of the superlattice is

$$H=\mu_0 B_z(g_e S_z+g_h \Sigma_z),$$

where μ_0 is the Bohr magneton, $S_z=1/2$ is the spin quantum number of the electron, and $\Sigma_z=3/2$ is the effective quantum number of the heavy hole.

From this it is easy to find the effective g-factor of the hole in the exciton. (The electron g-factor is taken to be constant and independent of the magnetic field at $g_e=-0.4$.)

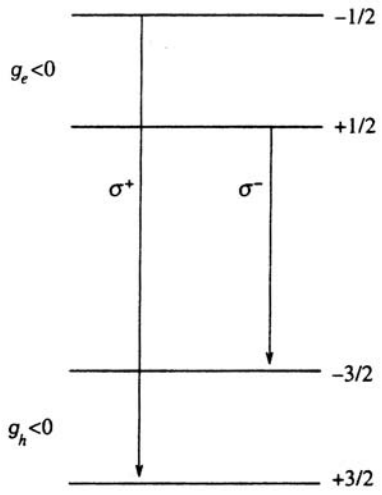

FIG. 8. Optical transition scheme for heavy hole excitons in a magnetic field.

Thus, the magnitude of the splitting implies that the effective g-factor of the hole in the $2sHH$ exciton is $g_h=-2.3$ in a 10 T field. In a $3sHH$ exciton, the magnitude of the g-factor is somewhat greater, with $g_h=-3.2$. With a reduction in the superlattice barrier width (superlattices 80/30/80 and 80/20/80), which is accompanied by weakened confinement, the magnitude of the doublet splitting and the corresponding hole g-factor of the magnetoexcitons decrease, in qualitative agreement with the concept of mixing of magnetoexciton states.[25]

b. The behavior of the intensity of the excited magnetoexciton states (in particular, the $3sHH$ state) as a function of magnetic field was also surprising. This effect shows up most clearly in luminescence excitation spectra of the 80/30/80 superlattice (see Fig. 6). This effect is absent in the 80/50/80 and 80/100/80 superlattices with narrow minibands. Evidently, at low fields the oscillator strength of the $3sHH$ state, like that of the other more excited states, is low. In magnetic fields of 8 T or above, the intensity of this state in the spectrum begins to rise dramatically, so that at 9 T it has risen by more than an order of magnitude. With a subsequent increase in the magnetic field, the intensity of the $3sHH$ magnetoexciton decreases and returns to its original value at 13 T. The sharp nonmonotonic variation in the intensity of the $3sHH$ exciton with magnetic field shows that this state interacts resonantly in the overlap region with another dipole-allowed state.

In this same spectral region, resonant intensity behavior emerges for the $2sHH$ exciton as well, but naturally in another range of magnetic fields. Calculations of the electron–hole size-quantization spectrum of the 80/30/80 superlattice show that in the spectral region where resonant behavior of the intensity of the $3sHH$ and $2sHH$ excitons is observed, there exist an M_1 saddle point of the electron–hole light hole miniband $[e(1)-LH(1)]$ and a minimum of the next size-quantization band for heavy holes $[e(1)-HH(2)]$. Even in the absence of a magnetic field, appreciable resonant absorption can be seen in this region, which is probably associated with these extrema of the minibands (see Fig. 6). The experimental facts discussed here are a qualitative indicator of the resonant character of the interaction of the excited magnetoexciton states with magnetoexciton states corresponding to these extrema of the minibands. However, the nature of this interaction is unclear, and a quantitative description of this phenomenon remains a task for the future.

6. CONCLUSION

We have studied spatially indirect exciton states with heavy and light holes in terms of luminescence and luminescence excitation spectra in GaAs/Al_xGa_{1-x}As superlattices in the flat band regime. It has been shown that the optimum conditions for observing indirect excitons are realized when the Coulomb energy in an exciton is of the order of the width of the electron–hole miniband in the superlattice. Significantly, observing these states does not require breaking of inversion symmetry. A strong magnetic field perpendicular to the heterolayers of the superlattice, which enhances exciton Coulomb bonding, will transform a system of interacting quantum wells to weaker bonding. Spatially indirect excitons in superlattices are analogous to the first order Wannier–Stark localization of excitons in superlattices with oblique bands; however, in the present case the localization of indirect excitons originates in the Coulomb interaction.

The authors thank K. Sørensen for preparing the superlattices.

This work was supported by the Russian Fund for Fundamental Research (Project 96-02-17535), the Scientific–Technical Program on the Physics of Solid State Nanostructures, and the INTAS (Grant 576i95). The authors express their sincere gratitude to these foundations.

[1] H. Chu and Y.-C. Chang, Phys. Rev. B **36**, 2946 (1987).
[2] A. Chomette, B. Lambert, B. Deveaud, F. Clerot, A. Regreny, and G. Bastard, Europhys. Lett. **4**, 461 (1989).
[3] H. Chu and Y.-C. Chang, Phys. Rev. B **39**, 10861 (1989).
[4] D. M. Whittaker, Phys. Rev. B **41**, 3238 (1990).
[5] M. M. Digam and J. E. Sipe, Phys. Rev. B **43**, 4097 (1991).
[6] P. M. Young, H. Ehrenreich, P. M. Hui, and N. F. Johnson, J. Appl. Phys. **74**, 7369 (1993).
[7] B. Deveaud, A. Chomette, F. Clerot, A. Regrey, J. C. Maan, R. Romestain, and G. Bastard, Phys. Rev. B **40**, 5082 (1989).
[8] K. Fujiwara, K. Kawashima, T. Yamamoto, N. Sano, R. Gingolani, H. T. Grahn, and K. Ploog, Phys. Rev. B **49**, 1809 (1994).
[9] K. Fujiwara, K. Kawashima, T. Yamamoto, and K. Ploog, Solid State Electron. **37**, 889 (1994).
[10] E. Mendez, F. Agulli-Rueda, and F. Hong, Phys. Rev. Lett. **60**, 3426 (1988).
[11] P. Voisin, J. Bleuse, C. Bouche, S. Gaillard, C. Alibert, and A. Regreny, Phys. Rev. Lett. **61**, 1639 (1988).
[12] K. Leo, P. Bolivar, F. Bruggeman, and R. Schwedler, Solid State Commun. **84**, 943 (1992).
[13] H. Roskos, M. C. Nuss, J. Shah, K. Leo, D. A. B. Miller, A. M. Fox, S. Schmitt-Rink, and K. Kohler, Phys. Rev. Lett. **68**, 2216 (1992).
[14] D. Birkedal, K. El Sayed, G. Sanders, V. G. Lyssenko, C. Stanton, and J. M. Hvam, Nuovo Cimento D **17**, 1359 (1995).
[15] M. M. Dignam and J. E. Sipe, Phys. Rev. B **43**, 4084 (1991).
[16] C. W. Bryant, Phys. Rev. B **47**, 1683 (1993).
[17] T. Westgaard, Z. X. Zhao, B. O. Fimland, K. Johanessen, and J. Johnsen, Phys. Rev. B **45**, 1784 (1992).
[18] L. V. Butov, A. Zrenner, G. Abstreiter, A. V. Petinova, and K. Eberl, Phys. Rev. B **52**, 12153 (1995).
[19] M. Bayer, V. B. Timofeev, E. Faller, T. Gutbrod, and A. Forchel, Phys. Rev. B **54**, 8799 (1996).
[20] A. V. Dzubenko and A. L. Yablonskii, Phys. Rev. B **53**, 16335 (1996).

[21] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterojunctions*, Les Editions de Physique, Les Ulli, France (1988).
[22] D. Birkedal, K. El Sayed, G. Sanders, C. Spiegleberg, V. G. Lyssenko, C. Stanton, J. M. Hvam, V. B. Timofeev, and M. Bayer, Phys. Rev. B **54**, 10316 (1996).
[23] B. Deveaud, A. Chomette, A. Regreny, and F. Clerot, Europhys. Lett. **11**, 3657 (1990).
[24] M. J. Snelling, E. Blackwood, C. J. M. McDonagh *et al.*, Phys. Rev. B **45**, 3922 (1992).
[25] G. E. W. Bauer and T. Ando, Phys. Rev. B **37**, 3130 (1988); Phys. Rev. B **38**, 6015 (1988).
[26] V. B. Timofeev, M. Bayer, A. Forchel, and M. Potemski, JETP Lett. **64**, 57 (1996).

Translated by D. H. McNeill

Supercooling in a system of two superconductors

I. N. Khlyustikov

P. L. Kapitsa Institute for Physical Problems, Russian Academy of Sciences, 117334 Moscow, Russia; Institute for Problems of Microelectronic Technology and Superpure Materials, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

(Submitted 29 April 1997)
Zh. Éksp. Teor. Fiz. **112**, 1119–1131 (September 1997)

Supercooling in the transition of a type I superconductor to the superconducting state in contact with another superconductor whose critical temperature is higher has been measured. Using aluminum as a test material, it has been demonstrated that at temperatures below the critical temperature T_c and magnetic fields below the critical field $H_c(T)$, aluminum remains in a metastable normal state, in spite of its contact with another superconductor. This means that it is not possible to generate a thermodynamic instability in a superconductor's electronic system through the ''proximity effect'' with another superconductor whose critical temperature is higher. This experimental observation demonstrates a radical difference between surface superconductivity, which certainly generates instability in normal electronic states, and superconductivity induced by the proximity effect near a junction with another superconductor.
 [S1063-7761(97)02309-3]

1. INTRODUCTION

It is generally accepted that the superconducting phase propagates from one superconductor to another in the same manner as superconductivity emerging at the surface propogates throughout the volume. This opinion is based on the fact that in either case, the order parameter is described by a monotonic function of position that relatively slowly decays in the normal-phase volume over a distance approximately equal to the correlation length.

As for surface superconductivity, it is common knowledge that if the superconducting state has a lower thermodynamic potential, propagation of the superconducting phase throughout the superconductor volume is not impeded by any potential barriers. In the case of type I superconductors, this means that supercooling below the critical field line of surface superconductivity defined by the function $H_{c3}(T)$ is impossible. A discussion relevant to this statement can be found, e.g., in Ref. 1, and highly convincing experimental evidence is given by Feder and McLachlan.[2]

Now let us discuss two different superconductors in contact with one another. In this case, the so-called proximity effect occurs in the contact zone, i.e., electrons and Cooper pairs penetrate from one metal to another. A vast number of experimental and theoretical studies have been dedicated to the proximity effect. Structures composed of a superconductor and a normal metal have been studied especially thoroughly. Among the topics discussed is whether the region of supercooled metastable states could be suppressed by the proximity effect. This issue was raised in searching for superconductivity of noble metals and materials with the very lowest superconducting transition temperatures. Indeed, if a material is a type I superconductor, even the presence of a low residual magnetic field in the cryostat can be higher than $H_{c3}(T)$, and the superconducting state in question will not be detected. Buhrman and Halperin[3] claim that supercooled states of a superconductor cannot be realized if it is in contact with another superconductor with a higher T_c.

Doubts that the proximity effect could suppress supercooled superconductor states were expressed after studies of a twinning plane.[4,5] These works demonstrated that a twinning plane is a special metal with a superconducting transition temperature different from the critical temperature of the crystal volume surrounding the twinning plane. The superconductivity of twinning planes also has a different set of critical magnetic fields.

A typical feature of the effect of an isolated twinning plane on surrounding crystallites is that the region of metastable supercooled states is not eliminated completely, although it is much smaller than in the case of a single crystal. In the notation of Ref. 5, supercooled states occupy on the $H-T$ diagram the region between the curves $H_c(T)$ and $H^*(T)$, whereas in single crystals they reside between the curves $H_c(T)$ and $H_{c3}(T)$ $(H_c(T)>H^*(T)>H_{c3}(T))$. All of these curves intersect the temperature axis at T_{c0}, i.e., the critical temperature of bulk superconductivity.

It is common knowledge that twinning is one of the mechanisms of plastic deformation in all crystal structures, except those with the close-packed face-centered cubic (FCC) lattice.[6] In fabrication of bimetallic specimens, it is virtually impossible to avoid plastic deformation (if only, because of the difference between thermal expansion coefficients). Therefore, in order to check whether superconductivity of another metal can serve as a seed in some superconductor, a metal with FCC structure must be used as a test superconductor with a lower transition temperature.

This paper reports on measurements of parameters of the region of supercooled states in aluminum samples under conditions when the aluminum is in contact with another superconductor and without such a contact. As a control, similar measurements were performed on indium in contact with tin.

2. TESTED SYSTEMS

Presently four superconductors with an FCC lattice are known. There are indications that rhodium is also a super-

TABLE I.

Element		T_c, K	$H_c(0)$, Oe	Reference
Lead	Pb	7.1999	803.4	[7]
Thorium	Th	1.374	162	[8]
Aluminum	Al	1.1796	104.9	[7]
Iridium	Ir	0.100	20.1	[7]
Rhodium	Rh	0.002(?)	—	[8]

conductor. Table I lists critical temperatures and critical magnetic fields at zero temperature for these metals. Note that the critical temperatures of lead, aluminum, and iridium are recommended as reference temperatures.

In the context of the problem to be solved, lead cannot be used because its Ginzburg–Landau parameter κ is too large. In lead, the critical field $H_{c3}(T)$ of surface superconductivity is higher than $H_c(T)$, so it has no supercooled states. Moreover, lead is not coated with an oxide film impervious to atmospheric oxygen, so measurements of the surface superconductivity critical fields depend on the sample history. Thorium is inconvenient because of its radioactivity. Iridium and rhodium have critical temperatures unattainable in a ^{4}He cryostat. For these reasons, the only suitable superconductor for the planned experiment was aluminum, so its supercooled states have been investigated.

Now let us discuss requirements on the second superconductor of the system. Firstly, its critical temperature should be higher than that for aluminum. Secondly, it is desirable that the mutual solubility of aluminum and the second metal be minimal, otherwise a diffusion layer with variable composition could affect experimental results. Compounds of two metals with intermediate compositions like Al_xMe_y can have a similar effect. Finally, one must have some test of whether an oxide film separating two superconductors is present. It is obvious that a sufficiently thick insulating film can prevent penetration of the superconducting phase from one superconductor to another.

In the reported work, an attempt was made to satisfy all these requirements by two methods. The first of them is fabrication of a mechanical (or electrical) contact between two metals immersed in liquid helium. In this case, diffusion between metals or emergence of any intermetallic compounds is out of question. In order to destroy oxide films, a force generating a plastic deformation to a level of up to $\approx 5\%$ could be applied. Furthermore, when an electric potential difference is applied to the interface, electric breakdown and destruction of an oxide film occur. The potential difference was applied by connecting a charged capacitor to the structure and discharging it through the interface between the two metals. This technique was used in studies of the aluminum–tin and aluminum–aluminum structures, and also in control experiments with indium–tin interfaces.

Elimination of the oxide film is necessary in such a process as soldering. When two metals are soldered, oxide films are eliminated through a reaction with a flux, and the imposed metal (soldering metal or alloy) is in the liquid state. The second technique for fabrication of bimetallic systems was deposition of one metal in the liquid state on the surface of another metal.

Among metals with a melting temperature lower than that of aluminum, tin satisfies all the requirements on the second superconductor. Its superconducting transition temperature $T_c \approx 3.72$ K. The solubility in the Al–Sn system at the tin melting temperature is negligible, and no intermetallic compounds are formed.[9,10] The lack of an oxide film was tested by wetting aluminum with liquid tin. It is known that liquid tin does not wet an aluminum surface coated with an oxide film, but easily spreads over an aluminum surface free from the oxide film. As for other low-melting superconductors, mercury amalgamates aluminum, i.e., it forms chemical compounds with it, whereas thallium, indium, and lead do not wet aluminum even when it is free from the oxide film.

When tin was deposited on aluminum, oleic acid was used as a flux. Since the natural oxide film on the aluminum surface is very resistant to oleic acid, a chemically polished[11] single-crystal aluminum plate was preliminarily etched in potassium hydroxide (KOH). Immediately after etching, without washing, the aluminum plate was placed in the acid and heated to the tin melting temperature. A piece of tin placed on the aluminum surface melted and spread over the surface. Without etching in the alkali, the aluminum surface was coated with too strong an oxide film, and tin could not wet the surface (a piece of tin sat on the surface in the form of a ball-shaped drop of liquid metal). After fabrication of the bimetallic plate, samples with required dimensions of $1\times1\times1$ mm were cut from it by an electric-arc machine. The thickness of the tin layer on one face of the resulting cube was about 0.15 mm. The layers damaged by the electric-arc cutter remained on the cube surface.

Of high-melting superconductors, we used tantalum, with $T_c \approx 4.46$ K. The fabrication technique was the same as in manufacture of evaporators used in coating surfaces with aluminum in vacuum. A tantalum wire 0.3 mm in diameter with an aluminum strip wound around it was placed horizontally in a volume evacuated by an oil diffusion pump and heated by an electric current. In the process of heating, first aluminum was melted, then its oxide film was destroyed, and the drop of liquid aluminum took on a ball-like shape. After a further increase in temperature, the aluminum drop wetted the wire, probably when the tantalum oxide film was also destroyed so that it passed through the center of the drop of liquid aluminum. It seems improbable that significant quantities of Al_3Ta compound could be produced, since the dissolution of tantalum in liquid aluminum is very slow.[10] Moreover, when aluminum was in the liquid state, its entire volume was intensely stirred by thermal flows from the current-heated tantalum wire, i.e., the small impurity of Al_3Ta should be uniformly distributed over the entire aluminum volume. All the stages of this process could be visually observed if the vacuum hood was transparent. The resulting sample was a piece of tantalum wire about 10 mm long passing through the center of an aluminum ball with a diameter of ≈ 1 mm. Naturally, in this case formation of an aluminum single-crystal was out of question.

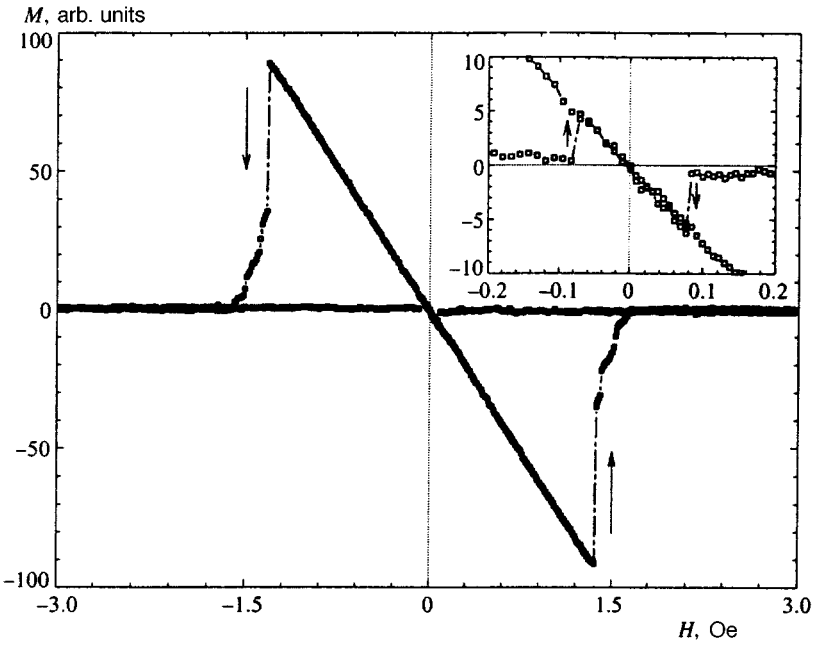

FIG. 1. Magnetic moment of an aluminum sample versus magnetic field around the transition to the superconducting state. The insert in the upper right quarter shows on an expanded scale the sections of the curves around the origin. The arrows indicate curve sections that were recorded when the field was scanned in a specific direction and jumps in the magnetic moment due to superheating and supercooling. $H_c = 1.61$ Oe, $H_{sc} = H_{c3} = 0.055$ Oe, $T \approx T_{c0} - 0.01$ K ≈ 1.17 K.

3. EXPERIMENTAL TECHNIQUES

Experimental records were curves of the sample magnetic moment measured using a SQUID magnetometer[12] versus magnetic field at constant temperature.

The magnetometer was equipped with a system of thermal shields made of copper foil and foam-plastic plates, and it was located inside a glass cryostat. The inside diameter of the helium Dewar was 80 mm. The consumption of liquid helium after its transition to the superfluid state was about 50 cm^3/h. Given this consumption of liquid helium, one can lower the temperature to that of the aluminum superconducting transition (and even lower) by pumping ^{4}He vapor if the pumping rate of the pumping station at a pressure of about 0.1 Torr is 100–150 liters/s. The vacuum pipeline connecting the cryostat to the pumping station had a bottleneck diameter of 50 mm. The pressure drop across the pipeline was 0.1–0.2 Torr.

The minimal temperature in these experiments measured using the superconducting transition temperature in an aluminum single crystal was about 1.1 K. For convenience, the pressure in the cryostat cap was measured using a McLeod manometer.

In order to perform the described experiments, new equipment for the magnetometer was designed and fabricated. First, an opportunity was created for multiple contacts between two pieces of metal to form the studied system during one low-temperature experiment. Relatively rough displacements in the range between 1 μm and 1 mm were produced mechanically from outside the cryostat using a fine-thread screw in the helium volume and a system of levers. Fine displacements were performed by a package of ten PZT disks with a diameter of 5 mm and thickness of 0.5 mm. The contact between pieces of metals was detected and its resistance was measured using a four-terminal configuration. The mechanical strength of the facility allowed one to apply a force sufficient for plastic deformation of aluminum to a level of 5–10%. The dimension of tested superconducting pieces could range down to ≈ 0.5 mm.

A second device allowed one to conduct measurements in different alignments of the sample with respect to an external magnetic field in the range of $\pm 180°$. The mechanical drive made it possible to change the alignment with a minimal step of $\approx 10^{-2}$ rad. This device was used in measurements of samples with plane, well-defined interfaces between two metals. Typical dimensions of samples that could be tested using this device were about 1 mm.

4. EXPERIMENTAL RESULTS

Before the beginning of the experiments described in this paper, it seemed that results and conclusions by Buhrman and McLachlan,[3] who studied the tungsten–titanium system, should also apply to the aluminum systems. For this reason, a sophisticated device for driving the superconductors to contact inside the cryostat was fabricated. It was assumed that, depending on the contact resistance between aluminum and another superconductor, the critical field of aluminum supercooling should vary.

Figure 1 shows a record of the aluminum single crystal magnetic moment versus magnetic field at a temperature below its transition temperature. It demonstrates all features typical of type I superconductors in magnetic field.

In fact, the graph shows two records. One of them was made by scanning the field from the left to the right, and it shows the field dependence of the aluminum magnetic moment in the range between a relatively large ''negative'' field (larger than $H_c(T)$ in absolute value) and a similar ''positive'' field. Another record was made by scanning the field in the opposite direction. These two curves coincide over three ranges. First (we have two such intervals), when aluminum is in the normal state, i.e., when $|H| > H_c(T)$. In this case $H_c = 1.61$ Oe, which was derived from the position of the

cusp in $M(H)$. The sample temperature derived from measurements of H_c was 1.17 K, which is about 0.01 K below T_{c0}.

Another region where the two curves coincide is $|H|<H_{sc}(T)$. The measured supercooling field H_{sc} is 0.055 Oe, which is about $H_c/30$. Given so large a value of the supercooling, one can assume that $H_{sc}=H_{c3}$. Hence, using the relation $H_{c3}=1.69\kappa H_c$, we obtain an estimate $\kappa_{\mathrm{Al}}=0.0143$, which is close to the value reported in the literature.

The difference between sample magnetic moments when the magnetic field is scanned up and down in the regions $H_{c3}(T)<|H|<H_c(T)$ results from the fact that the studied phase transition is of the first kind. When the field is reduced from large values, the sample remains in a metastable normal state and has a magnetic moment close to zero. At lower fields, $|H|<H_{c3}(T)$, the only stable state is the superconducting state with magnetic susceptibility $\chi=-1/4\pi$. As the field drops and reaches $H=H_{c3}$, the magnetic moment abruptly increases from zero to a finite value. When the absolute value of the external field gradually increases from zero, the superconducting state characterized by the complete Meissner effect remains stable up to 1.37 Oe in this specific case.

The superheating in increasing magnetic field shows up in the range 1.35 Oe$>|H|>$1.07 Oe$\approx(2/3)H_c$. Over this range, the sample can be either in the superheated state (as in Fig. 1) or in an intermediate state with a positive differential magnetic susceptibility. In the example being discussed, aluminum is in the intermediate state in the field range 1.35 Oe$<|H|<H_c$.

Then the suitability of aluminum for experiments with bimetallic systems was tested. To this end, an additional sample chemically identical to the first one but strongly deformed at room temperature was mounted in the magnetometer. The samples were insulated from one another, and in order to create identical conditions for them (i.e., equal magnetic fields applied to them), the separation between them was minimized. Measurements demonstrated that both aluminum pieces, the single crystal and the deformed sample, had equal supercooling fields. On the curves of $M(H)$, only one jump of magnetic moment corresponding to the transition from the normal to superconducting state in both samples was detected, although their manufacturing techniques were different. In other words, a situation when one piece was in the superconducting state and the other in the metastable normal state has not been observed.

After these tests, structures composed of different superconductors, namely aluminum–tin and aluminum–tantalum, could be investigated.

In the device that allowed one to bring to contact and to separate two metal pieces, an aluminum single crystal and a piece of tin were mounted. Initially they were at a distance of about 0.2 mm from one another. The curves of magnetic moment versus magnetic field due to the aluminum transition to the superconducting state are shown in Fig. 1. As concerns tin, note the following feature of this material. When the magnetic field is scanned over the range between -5 Oe and $+5$ Oe, nonlinear features in the curve of $M(H)$ due to the superconducting transition can be observed only around its critical temperature, which is $T_c\approx3.72$ K, as was noted above. At temperatures close to the aluminum critical temperature, all critical fields of tin are considerably higher than the fields used in the experiments. Therefore, in the records shown on all graphs of this paper the piece of tin was always in the homogeneous superconducting state. Thus, the magnetometer signal due to the tin sample was linear and proportional to the external magnetic field, i.e., $M=(-1/4\pi)H$. In the graphs given in this paper, a linear part of the $M(H)$ curve was extracted. The pictures present only the signal due to the aluminum superconducting transition. Furthermore, it was assumed that for $H>H_c$ the magnetic moment was constant.

A completely unexpected result was obtained when aluminum and tin were brought into close contact. It turned out that irrespective of whether there was contact between aluminum and tin or not, the curves of the magnetic moment versus magnetic field due to the superconducting transition were identical. The electric resistance of the contact varied from infinity to less than 10^{-3} Ω, and in all cases the fields H_c and H_{c3} remained unchanged. Thus, the contact between the pieces of aluminum and tin did not alter the range of the aluminum supercooled normal state.

The situation did not change after a force sufficient to generate a 5% plastic deformation of tin and aluminum was applied in order to produce a tighter contact. The magnitude of plastic deformation was estimated by measuring sample dimensions after removing them from the cryostat.

A similar result was produced by an experiment in which oxide films in the contact area were destroyed by discharging a capacitor.

Since doubt about the presence of oxide films in mechanical contacts between metals and their effect on measurements could not be dispelled completely, experiments were continued with samples fabricated using alternative techniques described in the previous section.

Figure 2 shows as an example curves of the aluminum superconducting transition obtained at a temperature of 1.165 K in samples fabricated by liquid tin deposition on aluminum.

Although the shapes of these curves are not quite ''classical,'' unlike those in Fig. 1, the result is essentially the same, namely, the effect of supercooling is also observed in this case. Both the critical field H_c and the supercooling field H_{sc} can be derived from these curves. In Fig. 2 they are 2.28 Oe and 0.68 Oe, respectively. Naturally, $H_c>H_{sc}$. The difference between the supercooling critical field H_{sc} and H_{c3}, which is a characteristic of aluminum, is most probably due to structural damage in the process of electric-erosion treatment, although it might also be due to the effect of tin on aluminum.

The instrumental effect resulting in the difference between the baseline magnetic moments on the right and on the left of Fig. 2 and also in the difference between jumps in the moment at $\pm H_{sc}$ should be elucidated. Using a SQUID as a null detector, the input signal could be canceled by a negative feedback signal so that the SQUID could remain in the

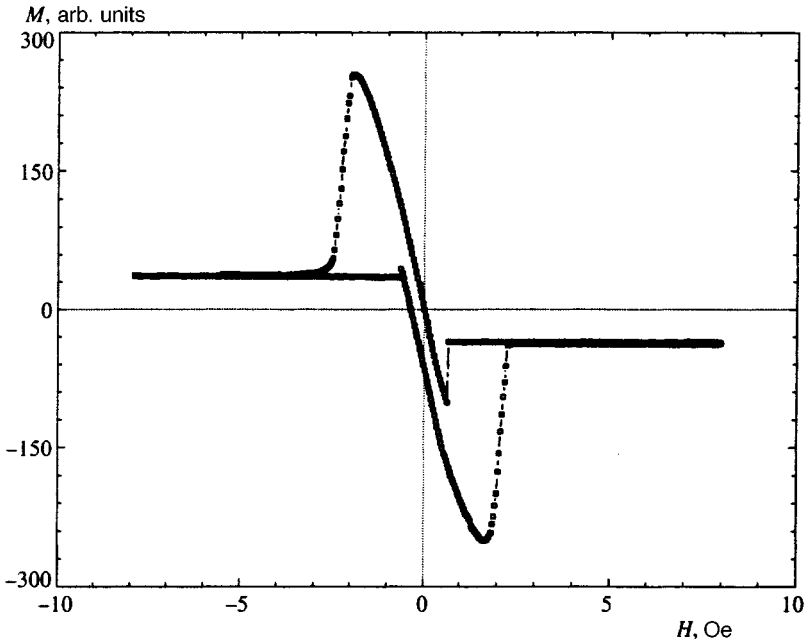

FIG. 2. Magnetic moment of an aluminum sample with an almost cubic shape with tin deposited on one face, as a function of magnetic field; $H_c = 2.28$ Oe, $H_{sc} = 0.68$ Oe, $T \approx 1.165$ K.

same quantum state. Since the response rate of feedback circuits is always finite and sometimes insufficient, the feedback circuit cannot compensate for the input signal when its amplitude is large and it changes at a high rate. In such a situation, the SQUID may switch to another quantum state, and this shows up as a change in the baseline level. These issues and techniques for recovering the signal without distortions caused by limitations of the feedback circuits were discussed in detail elsewhere.[12]

Measurements performed at various sample alignments in a magnetic field did not show any appreciable changes in H_c and H_{sc}. Only the shapes of the magnetization curves changed slightly, which was quite natural since the samples were not ball-shaped.

Figure 3 shows similar curves for aluminum deposited on tantalum. One can see that in this case the regions of supercooled states are also clearly defined, since H_c and H_{sc} are 6.5 Oe and 5.0 Oe, respectively. The experiment was conducted at a temperature of 1.14 K.

Now let us consider the control experiment, which can at least partly clarify results reported by Buhrman and Halperin.[3] In the control experiment, the device for bringing the samples into mechanical contact many times at liquid-helium temperature was used again, and an indium single crystal was mounted instead of aluminum. The superconducting transition temperature in indium is about 3.4 K. The second superconductor was tin.

Indium has a tetragonal crystal lattice, and twins are eas-

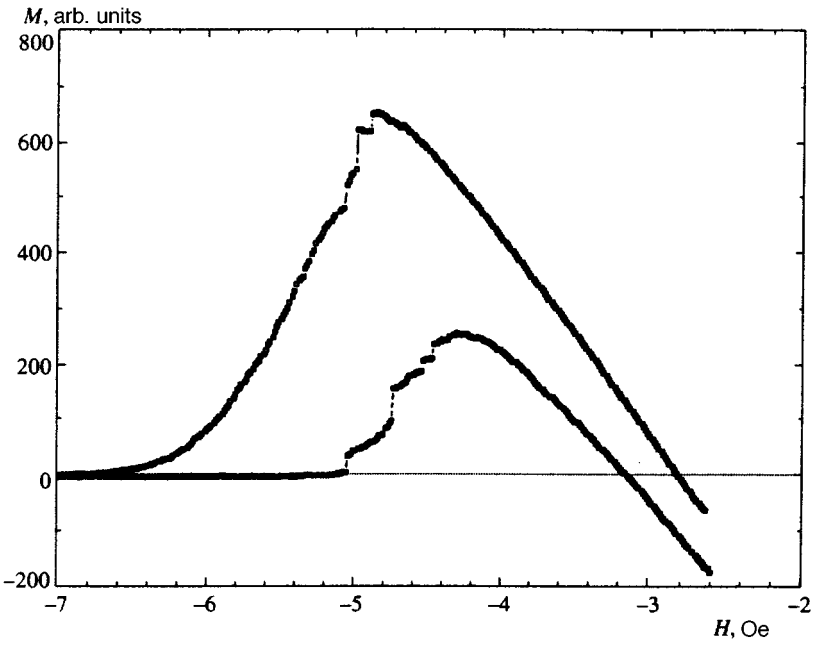

FIG. 3. Magnetic moment of aluminum versus magnetic field around the superconducting transition. Aluminum was deposited in the form of a liquid drop on a tantalum wire; $H_c = 6.5$ Oe, $H_{sc} = 5.0$ Oe, $T \approx 1.14$ K.

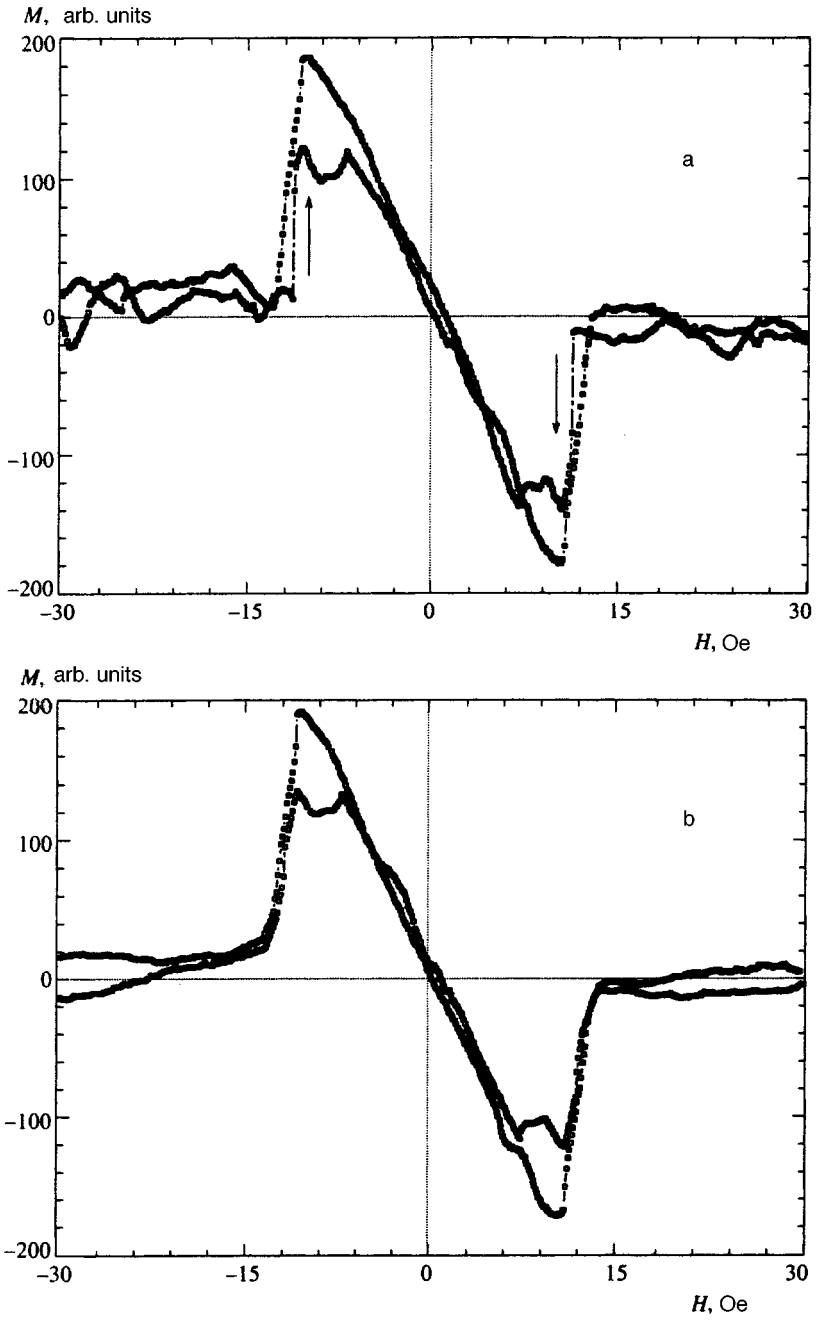

FIG. 4. Magnetic moment of indium as a function of magnetic field around the superconducting transition: a) before plastic deformation at liquid-helium temperatures; the arrows indicate jumps in the magnetic moment at the supercooling field, $H_c = 12.9$ Oe, $H_{sc} = 11.3$ Oe, $T = 3.3$ K; b) after plastic deformation. No supercooling could be detected.

ily generated in the [101] plane.[6] Therefore it is very difficult to fabricate an indium sample that demonstrates supercooling corresponding to its characteristic value of the Ginzburg–Landau parameter κ. Indium supercooling to the fields H_{c2} and H_{c3} was reported, for example, in Ref. 2. In order to detect a small change in the indium supercooling field corresponding to the superconducting transition, it is sufficient to anneal the sample at room temperature.

Figure 4a shows experimental curves of the indium sample magnetic moment versus magnetic field recorded at a temperature of about 3.3 K. The critical magnetic field $H_c = 12.9$ Oe, and the supercooling field $H_{sc} = 11.3$ Oe.

Contact between indium and tin under a force below the plastic deformation threshold did not change the curves $M(H)$ describing the superconducting transitions down to contact resistances below $10^{-3}\ \Omega$. Multiple contacts and separations between the two pieces did not affect the indium magnetization curves.

The situation changed radically when a force sufficient for plastic deformation was applied. The supercooling region disappeared, as shown in Fig. 4b. Subsequent separation of the samples did not lead to restoration of the supercooling effect.

A similar result was obtained in experiments with plastic deformation generated in indium by a quartz tip.

5. CONCLUSIONS

The reported experiments indicate that aluminum is a good test material for studies of superconductivity in bime-

tallic systems. The requirements on crystallographic perfection of aluminum samples in experiments similar to those described above are very low. The parameters of the supercooling region in plastically deformed samples are the same as in single crystals, at least in the neighborhood of the temperature corresponding to the superconducting transition.

The most important result of the discussed experiments is that there are radical differences between the surface superconductivity and superconductivity induced by another superconductor. The main difference is that the superconductivity due to the proximity effect of another superconductor does not cause a thermodynamic instability in metastable normal states. In other words, the superconductivity induced by another superconductor does not act as a seed with supercritical parameters from which the superconducting phase propagates throughout the superconductor volume, and this results in a type I phase transition.

Certainly, there are no grounds to assert that one superconductor does not affect another. Moreover, experiments with twins[4] clearly indicate that such an effect actually occurs. This effect, however, is such that there is a finite energy barrier preventing growth of the superconducting phase in the superconductor bulk.

In terms of the Ginzburg–Landau theory, a superconductor in close contact with a normal metal generated a certain value of the order parameter Ψ_0 on the metal surface. The coordinate dependence of the absolute value of the order parameter, $|\Psi(x)|$, is well known: it decays with distance from the interface. Such behavior is also typical of surface superconductivity. If the effect of the boundary values of $d\Psi/dx$ on the phenomenon under consideration is neglected, the following differences can result. In the case of the surface superconductivity, the order parameter on the surface $|\Psi_0|$ is known to be close to unity, i.e., the value it has when the superconducting phase occupies the entire volume.[13] It is possible that in the case of the proximity effect (induced superconductivity), $|\Psi_0| \ll 1$. Even a small value of the order parameter is sufficient to screen external magnetic fields, i.e., to generate nondissipative currents in a certain layer near the metal surface. But since $|\Psi_0|$ is small, the contribution to the system energy from the induced superconductivity remains positive. Such a ''seed'' is characterized by parameters below the critical values and cannot cause absolute instability in metastable states.

Thus, there should be a certain critical value Ψ_0^* above which the induced superconductivity, like the surface superconductivity, acts as a seed with supercritical parameters. The question is, what is this critical value Ψ_0^*? In addition, there is a question of whether Ψ_0 due to the proximity effect can be increased. Otherwise, if Ψ_0 is always small, the region of induced superconductivity cannot act as a superconducting phase seed, and supercooled metastable states should persist.

The statement above applies only to the case when two different superconductors are in contact on a certain surface. Strictly speaking, only in this case does the comparison between surface and induced superconductivity make sense. A radically different situation occurs when a notable fraction of the superconductor volume contains a lot of inclusions of the other superconductor with higher critical characteristics—for example, twinning planes. No supercooling was observed in experiments.[14] Thus, the volume fraction with twins serves as a seed leading to instability of metastable states. But this effect is limited to superconductors in which a considerable concentration of twins can be generated by plastic deformation, i.e., materials with crystal lattices different from close-packed FCC. It is possible that grains of a superconductor with higher T_c in sufficient concentration can also destroy supercooled states.

In view of the experiments described above, the results reported by Buhrman and Halperin[3] can have a feasible interpretation completely different from that given by the authors. The reported results can be ascribed not to the proximity effect, but to the technique of sample fabrication. Tungsten and titanium were joined by spot welding, i.e., a very harsh procedure was used, which could not be done without plastic deformation. Hence, since the tungsten crystal lattice is not close-packed FCC but body-centered, mechanical twins could be generated near the welding spots, and they could lead to elimination of the supercooling effect. A test of this interpretation could be an experiment with a nonsuperconducting metal welded to tungsten.

The author would like to express gratitude to A. F. Andreev, V. A. Tulin, V. V. Lebedev, V. S. Edelman, N. V. Zavaritskiĭ, and E. V. Mininberg for helpful discussion of topic related to the reported work.

[1] P. G. de Gennes, *Superconductivity of Metals and Alloys*, W. A. Benjamin, New York–Amsterdam (1966).
[2] J. Feder and D. S. McLachlan, Phys. Rev. **177**, 763 (1969).
[3] R. A. Buhrman and W. P. Halperin, J. Low Temp. Phys. **16**, 409 (1974).
[4] I. N. Khlyustikov, Zh. Éksp. Teor. Fiz. **94**, 314 (1988) [Sov. Phys. JETP **67**, 607 (1988)].
[5] I. N. Khlyustikov and A. I. Buzdin, Adv. Phys. **36**, 271 (1987); I. N. Khlyustikov and A. I. Buzdin, Usp. Fiz. Nauk **155**, 47 (1988) [Sov. Phys. Usp. **31**, 409 (1988)].
[6] M. V. Klassen-Neklyudova, *Mechanical Twinning*, Izd. Akad. Nauk SSSR, Moscow (1960).
[7] M. Durieux, D. N. Astrov, W. R. G. Kemp, and C. A. Swenson, Metrologia **15**, 57 (1979).
[8] *Superconducting Materials*, [in Russian], E. M. Savitskiĭ, (ed.) Metallurgiya, Moscow (1976).
[9] W. B. Pearson, *A Handbook of Lattice Spacings and Structures of Metals and Alloys*, Pergamon Press (1958).
[10] V. S. Sinel'nikova, V. A. Podergin, and V. N. Rechkin, *Aluminides*, [in Russian], Naukova Dumka, Kiev (1965).
[11] W. Tegart, *Electric and Chemical Polishing of Metals* [Russian translation], Izd. Inostr. Lit., Moscow (1957).
[12] I. N. Khlyustikov, Prib. Tekh. Eksp. **2**, 167 (1993).
[13] D. Saint-James, G. Sarma, and E. J. Thomas, *Type II Superconductivity* Pergamon Press, Oxford (1969).
[14] I. N. Khlyustikov and M. S. Khaĭkin, JETP Lett. **38**, 224 (1983).

Translation was provided by the Russian Editorial office.

Arnold diffusion in large systems

B. V. Chirikov* and V. V. Vecheslavov

Budker Institute of Nuclear Physics, 630090, Novasibirsk, Russia
(Submitted 19 November 1996)
Zh. Éksp. Teor. Fiz. **112**, 1132–1146 (September 1997)

A new regime of Arnold diffusion in which the diffusion rate has a power-law dependence on the perturbation strength is studied theoretically and in numerical experiments. The theory developed predicts this new regime to be universal in the perturbation intermediate asymptotics, the width of the latter increasing with the dimensionality of the perturbation frequency space, particularly in large systems with many degrees of freedom. The results of numerical experiments agree satisfactorily with the theoretical estimates. © *1997 American Institute of Physics.* [S1063-7761(97)02409-8]

1. INTRODUCTION: UNIVERSAL NONLINEAR INSTABILITY

One of the most interesting phenomena in Hamiltonian dynamics is the so-called Arnold diffusion (AD), a distinctive universal instability of multidimensional nonlinear oscillations.[1,2] This global instability was predicted by Arnold;[3] its chaotic nature was discovered in Refs. 1, 4, and 5 and further studied in detail in Refs. 6–11, 14, 15, and 17.

First, following Ref. 17, we briefly recall the diffusion mechanism, which is related to the interaction of nonlinear resonances. Consider a general Hamiltonian describing multidimensional oscillations:

$$H(I,\theta,t)=H_0(I)+\varepsilon\sum_{n,m} V_{nm}(I)\exp(in\cdot\theta+itm\cdot\Omega), \tag{1.1}$$

where I, θ are N-dimensional vectors of the action–angle variables; Ω is the M-dimensional vector of the driving frequencies; n, m are integer vectors of dimensions N and M, respectively, and ε stands for a small perturbation parameter. The dot in expressions like $n\cdot\theta$ denotes the scalar product. Below we shall consider the simpler case of a completely integrable and nondegenerate unperturbed system whose Hamiltonian $H_0(I)$ depends on the full set of N actions only.

Hamiltonian (nondissipative) dynamics is always determined by resonances (see, e.g., Refs. 1 and 2) corresponding to particular terms in the perturbation (1.1). The condition for a primary resonance with unperturbed frequencies (1.3) is

$$\omega_{nm}\equiv n\cdot\omega(I)+m\cdot\Omega\approx 0. \tag{1.2}$$

In the case of linear oscillations all the frequencies are fixed as parameters of the system which is either in or off resonance independent of initial conditions. However, for nonlinear oscillations with the action-dependent frequencies

$$\omega(I)=\partial H_0(I)/\partial I, \tag{1.3}$$

condition (1.2) determines resonance surfaces (zones) in the phase space, that is, the system is always in resonance for some initial conditions. On the other hand, nonlinearity stabilizes the impact of a (sufficiently weak) perturbation, ensuring bounded oscillations even for resonant initial conditions. This is precisely due to non-isochronous oscillations (1.3). In one degree of freedom such a nonlinearity is necessary and sufficient to destroy the oscillation isochronism. The generalization of that for several degrees of freedom is the necessary condition for determinant to be nonzero everywhere,

$$\left|\frac{\partial^2 H_0}{\partial I^2}\right|\neq 0. \tag{1.4}$$

In this case the system is called nondegenerate. This allows, in particular, the transformation from action to frequency space. In the latter, the resonance structure is especially simple and transparent, as the resonant surfaces (1.2) become planes.

Another condition for the nonlinear stabilization is the requirement for the quadratic form associated with the matrix $\partial^2 H_0/\partial I^2$ to be sign-definite or, geometrically, for the surfaces $H_0(I)=\text{const}$ to be convex.[10] The latter condition is a weaker one as it may include higher polynomial forms. Both conditions are only sufficient.[10,11]

The above conditions also ensure the absence of strong instability ($\sim\varepsilon$), due to a quasilinear (isochronous) resonance,[1] especially when several (r) independent resonance conditions (1.2) are simultaneously satisfied. The latter is called multiple (r-fold) nonlinear resonance. However, a weak instability caused by nonresonant ($\omega_{nm}\neq 0$ for given initial conditions) terms in the perturbation series (1.1) is possible, and it is just the AD we are going to discuss in detail. Moreover, this weak instability is a typical phenomenon of nonlinear oscillations, since it occurs for almost any perturbation of a completely integrable system particularly one that is arbitrarily weak. The only restriction is the action space dimension d_a, which must be larger than that of the invariant torus ($d_a>d_t=1$).[3] The torus is an absolute barrier for the motion trajectory, which can only bypass it but never go through. For a driving perturbation ($M>0$ in Eq. (1.1)) the minimum number of degrees of freedom is, thus, $N_{\min}=2$, but in the conservative case ($M=0$) it is $N_{\min}=3$, since the trajectory is constrained to follow an energy surface.

Even these minimal restrictions are not absolute, since they apply to the strong nonlinearity (1.4) only when the effect of resonant perturbation is small ($\Delta I/I\sim\sqrt{\varepsilon}\ll 1$). In case of linear $H_0(I)$ (the harmonic oscillator) $N_{\min}$ is smaller by 1.[12]

At least three perturbation terms in the series (1.1) are necessary for AD. We shall call each of these terms a resonance (for the appropriate initial conditions of the motion). A single resonance retains the complete integrability of the unperturbed system. The interaction of even two resonances results in the formation of narrow chaotic layers around the unperturbed separatrices of both resonances,[13–15] but the chaotic motion remains confined within a small domain of the layer. Only the combined effect of at least two driving resonances gives rise to diffusion along the layer of the first, guiding, resonance if $N \geqslant N_{\min}$ holds (see Ref. 1 for details).

In the first approximation (1.2) the driving perturbation terms are nonresonant ($\omega_{nm} \neq 0$), but the final effect is due to the secondary resonances between the driving perturbation and the slow phase oscillation on the guiding resonance. This is a particular case of the general rule that all the long-term effects in nonlinear oscillations are due to some resonances. For the problem in question the principal parameter is the ratio

$$\lambda = \frac{|\omega_{nm}|}{\omega_g}, \tag{1.5}$$

where $\omega_g \sim (\varepsilon |V_g|)^{1/2}$ is the frequency of small phase oscillations at the center of the guiding resonance, and where V_g is the Fourier amplitude of the corresponding perturbation term. For a weak perturbation ($\varepsilon \to 0$) the parameter $\lambda \gg 1$ is big, and thus the effect of the driving resonances is a high-frequency one. In fact, this is equivalent to a low-frequency (adiabatic) perturbation. Hence we use the term inverse adiabaticity.[14] The symmetry between the standard and inverse adiabaticity is especially clear in a conservative system, i.e., for the interaction of coupling resonances. Indeed, in this case the resonant interaction results in energy exchange between the guiding and driving resonances. While for the former the perturbation is a high-frequency one (inverse adiabaticity), for the latter it is low-frequency (standard adiabaticity).

For an analytic perturbation the effect in both cases is exponentially small in the adiabaticity parameter λ (1.5), namely:[1,14]

$$D \sim e^{-\pi\lambda} \sim w_s^2, \tag{1.6}$$

where D is the local dimensionless diffusion rate in the action I within a chaotic layer and where $w_s \sim |\Delta H_0|/\varepsilon V_g$ stands for the dimensionless layer width (for a more accurate estimate see Ref. 14). Notice that the effect (1.6) is of a nonperturbative nature, since $\lambda \sim \varepsilon^{-1/2}$ (see Eq. (1.5)).

This is the simplest resonant mechanism of AD. In particular models the accuracy of such a three-resonance approximation was found to be within a factor of 2, provided that the perturbation is not too weak, i.e., the adiabaticity parameter λ is not very big[1] (see also Sec. 3 below).

As $\lambda \to \infty$ the higher-order resonances with large harmonics numbers $|n_i|$, $|m_j| \to \infty$ come into play. Even though their amplitudes drop exponentially $V_{nm} \sim \exp(-\sigma k)$, where $k = \Sigma |n_i| + \Sigma |m_j|$, the detunings $|\omega_{nm}|$ also rapidly decrease. The operative resonances which control the diffusion have been roughly identified in Refs. 1 and 15 by minimizing the expression

$$-\ln D \equiv E \sim k + \lambda(k) \gtrsim \lambda_0^{1/L} \tag{1.7}$$

with respect to k. Here $\lambda_0 = \omega_0/\omega_g$, ω_0 stands for a characteristic oscillation frequency, and the following diophantine estimate was used:

$$\omega_{nm} \sim \frac{\omega_0}{k^{L-1}}. \tag{1.8}$$

The most important parameter in Eq. (1.7),

$$L = N + M - r, \tag{1.9}$$

is the number of linearly independent (incommensurate) unperturbed frequencies on an r-fold resonance. We shall call L the resonance dimension (in frequency space). Actually, Eq. (1.9) gives the maximum dimension when all L independent frequencies contribute to the driving resonances, which may be termed the full resonances. There are also partial resonances which depend on a smaller number of frequencies $\underline{L} < L$. Even though there are only a few of the latter, they are crucially important for the new AD regime which is the main subject of this paper (Sec. 5).

The estimate (1.7), which represents another AD mechanism, seems to agree with numerical data.[7,14] On the other hand, Nekhoroshev rigorously proved[10] an upper bound of the form (1.7) but with a different exponent ($M = r = 0$):

$$L \leqslant L_N = \frac{(3N-1)N}{4} + 2. \tag{1.10}$$

Even for the minimum dimensions $N = 3$ this upper bound $L_{\max} = 8$ considerably exceeds the estimate (1.9): $L = 2(r = 1)$. The difference grows as $N \to \infty$. Even though this discrepancy is not a direct contradiction inasmuch as Eq. (1.10) is the upper bound, it constitutes a problem: what would be the origin of the difference between the two estimates?

Recently, this problem has been resolved by Lochak[11] who rigorously proved a more efficient Nekhoroshev-type estimate with the exponent (1.9) (for $M = 0$ but any r). The explanation is that Lochak assumed convexity of the unperturbed Hamiltonian $H_0(I)$ given above, whereas Nekhoroshev's proof holds under a weaker condition of the so-called steepness of H_0. From the physical point of view this difference appears to be insignificant. At least, we are not aware of any example of a steep but non-convex H_0.

Both the diffusion rate and the measure of the chaotic component ($\sim w_s$, see Eq. (1.6)) are exponentially small in the perturbation in the limit $\varepsilon \to 0$, hence the term KAM integrability[14] referring to the Kolmogorov-Arnold-Moser theory which proves the complete integrability for most initial conditions as $\varepsilon \to 0$. This partial integrability, or better, almost-integrability, is as good as the approximate adiabatic invariance. Notice, however, that the complementary set of initial conditions supporting AD—the so-called Arnold web—is everywhere dense, as is the set of all resonances (1.2), any one of which can be a guiding resonance. Also, the variation is exponentially slow in the action I only while the variation in oscillation constant (for the unperturbed motion) phase θ_0 is much faster, with a characteristic time of order

the inverse Lyapunov exponent, $\theta_0 \sim \omega_g / |\ln w_s| \sim T_w^{-1}$, where T_w is the oscillation period in the chaotic layer (see Eq. (2.2) below).

Both rigorous estimates are valid asymptotically, for sufficiently small ε only. For example, Lochak requires[11] ($L \gg 1$)

$$\varepsilon < \varepsilon_L \sim \left(\frac{\sigma^2}{L} \right)^{2L^2}, \tag{1.11}$$

where σ is some average decay rate of the perturbation amplitudes. This is very small perturbation, and the problem arises of estimating the diffusion rate in the intermediate asymptotic region: $\varepsilon_L \ll \varepsilon \ll 1$, or $1 \ll \lambda_0 \ll \lambda_L$. This problem was first addressed in Refs. 14, where a new regime of diffusion, called the fast Arnold diffusion (FAD), was conjectured from some preliminary results of numerical experiments. Two characteristics of the new regime as contrasted to the far-asymptotic AD (1.11) are as follows:

(i) the dependence of the diffusion rate on the adiabaticity (perturbation) parameter λ_0 (1.7) is a power law rather than exponential, and

(ii) the diffusion rate does not depend on the resonance dimension L, in particular, on the number of degrees of freedom N (cf. Eq. (1.7)).

Precisely this behavior has been observed in numerical experiments with another multidimensional model.[16] However, the authors of Ref. 16 have given a different interpretation of their numerical results. Instead, we tried to reconcile the same results with our new diffusion mechanism.[17] Unfortunately, both interpretations remain somewhat ambiguous because the perturbation in those numerical experiments was not sufficiently small to reach any asymptotic behavior where the theoretical estimates were expected to hold true. To resolve this ambiguity we continued numerical and theoretical studies with the same model but using a much weaker perturbation. In this paper we report on our first results and present their theoretical explanation.

2. MODEL AND NUMERICAL EXPERIMENTS

Following Refs. 16 and 17 we make use here of the same model with Hamiltonian

$$H(x,p,t) = \frac{|p|^2}{2} - K \sum_{i=1}^{N+1} \cos(x_{i+1} - x_i)\, \delta_1(t) \tag{2.1}$$

and periodic boundary conditions ($x_{N+2} = x_1$; $p_{N+2} = p_1$) where p, x are action-angle variables, $\delta_1(t)$ stands for the δ-function of period 1, and $K \to 0$ is small perturbation parameter. Notice that this model has N degrees of freedom due to the additional motion integral $\Sigma p_i = \text{const}$. The unperturbed frequencies $\omega_i = p_i$ are equal to the action variables, and the energy surfaces $H_0(p) = |p|^2/2 = \text{const}$ are spheres, and hence are strictly convex with unit determinant (1.4). The driving perturbation in the form of periodic «kicks» is not important for the diffusion but greatly simplifies numerical experiments as it allows the use of a (multidimensional) map rather than differential equations of motion.

Even though this model does not immediately represent by itself a physical system, it is very convenient for the studies of subtle nonlinear phenomena like AD. The resulting theory can, then, be applied to real physical problems, such as the stability of the Solar System[18] or of charged particles in magnetic fields in plasma devices, accelerators and colliding beams.[15,19]

In previous work the diffusion in multidimensional models like (2.1) was studied only down to $K \sim 0.1$.[16,9] For such perturbation levels and large N a considerable part of phase space becomes globally chaotic, which obscures the AD effect. Even though the combined action of AD and global diffusion is an interesting problem which is important for applications,[1,15] here we mainly wanted to understand the mechanism of AD itself. To this end we went down as far as to $K \sim 10^{-6}$ with up to $N = 15$ degrees of freedom. Realization of this program has required essential modification of the problem itself. This is because direct computation of the diffusion rate quickly becomes prohibitively slow as $K \to 0$, especially since a multiple computation precision is required for such a small K. To overcome this technical difficulty we have taken a different approach,[14] namely, computing the chaotic layer width w_s and recalculating the diffusion rate from a relation like (1.6). Of course, this makes sense for a model with $N \geqslant N_{\min}$ degrees of freedom (Sec. 1). In this way we have managed to reach (for another model) adiabaticity parameter values of $\lambda_0 \approx 50$ with an ordinary computer, as compared to $\lambda_0 \approx 10$ only for a direct diffusion calculation on a Cray supercomputer.[7] In the model (2.1) this would roughly correspond to $K \sim \lambda^{-2} \sim 4 \cdot 10^{-4}$ and 10^{-2}, respectively, and $N = 2$ only.

In the present work we go further, and give up the calculation of the diffusion rate altogether. Instead, we are studying numerically and developing the theory of the chaotic layer only. This proves sufficient to understand the mechanism of AD as well, since both are essentially determined by the same higher-order adiabaticity parameter (1.5) and the exponent in Eq. (1.7). Then, all we need in numerical experiments is to compute the oscillation period $T(w_s)$ inside the chaotic layer of a guiding resonance, and recalculate the layer width w_s using the simple relations[1]

$$\omega_g T_{\min} = \ln \frac{32}{w_s}, \quad \omega_g T_{\text{av}} = \ln \frac{32}{w_s} + 1, \tag{2.2}$$

where $T_{\min}$, T_{av} are the shortest and average periods, respectively. The two values are in a reasonable agreement, $\langle \ln(w_{\min}/w_{\text{av}}) \rangle = 0.31$, within the rms fluctuations $\Delta \ln(w_{\min}/w_{\text{av}}) = \pm 0.39$, and both underestimate the full layer width. This is because the diffusion at the layer edge is very slow, so that the 100 oscillation periods used in numerical experiments were insufficient to reveal the whole layer. A crude estimate[14] yields the expected correction factor of order 2. No such correction was introduced into the numerical data, but it will be discussed below in Sec. 3.

A primary coupling resonance $\omega_1 \approx \omega_2$ with phase oscillation frequency $\omega_g = \sqrt{2K}$ has been chosen as the guiding resonance. Correspondingly, $p_1 \approx p_2 \approx p_g$ while other p_i ($i = 3,...,N+1$) were taken at random (mod 2π). For the trajectory to be inside the layer the initial value of the guiding resonance phase was taken to be approximately $\psi_1 = x_1 - x_2 \approx \pi$. However, for small K the exact position of the

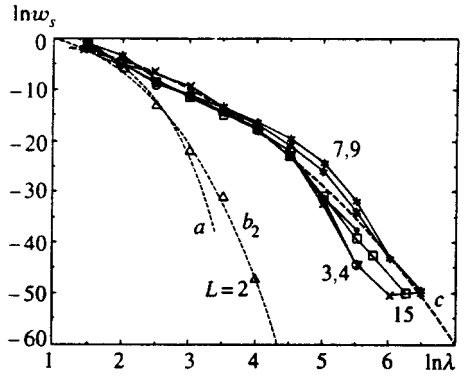

FIG. 1. Summary of numerical data for the model (2.1). Broken solid lines connecting various symbols show computed values of w_s as a function of the adiabaticity parameter $\lambda \equiv 1/\sqrt{K}$ and the resonance dimension $L=N$ indicated by the numbers. Dotted lines represent the theory: (a) small-λ limit, one fitting parameter, Eq. (3.5); (b_2) large-λ limit for $L=2$, two fitting parameters, Eq. (4.9); (c) intermediate asymptotics, three fitting parameters, Eq. (5.8).

layer had to be located numerically prior to computation of w_s by a special searching part of the code. The computation was performed for seven values $N=2$, 3, 4, 5, 7, 9, 15 with the same initial conditions for a single trajectory.

The results are summarized in Figs. 1 and 2. The lower bound of $w_s \sim 10^{-22}$ was determined by the computation precision (about 30 decimal places). The values of the principal model parameter—the number of independent unperturbed frequencies, or the resonance dimension $L=N+M-r=N$—are also indicated. Notice that under the particular conditions of the numerical experiments the resonance dimension is equal to the number of degrees of freedom of the model because the driving perturbation is periodic ($M=1$), and guiding resonance is simple ($r=1$).

The most striking feature of the empirical data is the qualitatively different behavior for $L=2$ which was observed already in Ref. 16. The rest of the data show no systematic dependence on L, but rather big fluctuations which rapidly increase with λ.

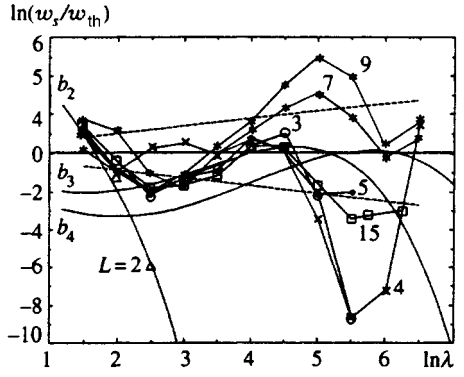

FIG. 2. The same data as in Fig. 1, with respect to the theoretical dependence $w_{th}(\lambda)$, Eq. (5.8) (curve c in Fig. 1). Thin solid curves $b_{\overline{L}}$ represent the first three members of the family $w_s(\lambda,\overline{L})$, Eq. (4.9) (cf. Fig. 3). Two dashed lines show rms w_s fluctuations (5.11).

3. SMALL-λ LIMIT: A SIMPLE DYNAMICAL THEORY

To lowest order in the small perturbation parameter K we can only consider the primary driving resonances which are explicitly present in the original Hamiltonian (2.1). Then the problem is very similar to that studied in Ref. 1, apart from a different expression for the kinetic energy. First, we transform the variables for the two degrees of freedom which determine the guiding resonance: x_1, x_2, p_1, $p_2 \to \psi_1$, ψ_2, I_1, I_2 where

$$\psi_1 = x_1 - x_2, \quad \psi_2 = x_1 + x_2,$$
$$p_1 = I_1 + I_2, \quad p_2 = I_2 - I_1. \tag{3.1}$$

In this approximation the momentum satisfies $I_2 \approx \psi_2$, and all $p_i \approx \dot{x}_i$ for $i \geqslant 3$ are constant and determine the frequencies of the driving resonances. The unperturbed motion on the separatrix of the guiding resonance is given by

$$\psi_1(t) = 4 \arctan(e^{\omega_g t}) - \pi, \tag{3.2}$$

where the frequency of the phase oscillation is $\omega_g = \sqrt{2K}$. As the interaction in the original Hamiltonian (2.1) is local, only the two degrees of freedom directly coupled to the guiding resonance contribute to the driving perturbation in the chaotic layer. The full set of driving resonances remains formally infinite because of the external perturbation $\delta_1(t)$ of frequency $\Omega = 2\pi$, but the effect of most of them is exponentially small due to the large detuning ω_{nm} (see Eqs. (1.5) and (1.6)). Consequently, one can retain a single driving resonance only with minimal detuning:

$$\omega_d = \min|p_g - p_d + s\Omega|, \tag{3.3}$$

where $p_d = p_3$, p_{N+1} and $s = 0. \pm 1$. In this approximation the Hamiltonian takes the form $H = H_0(I_1, \psi_1) + V(\psi_1, t)$, where

$$H_0 = I_1^2 - K \cos \psi_1, \quad V \approx -K \cos(\psi_1/2 - \omega_d t + \phi), \tag{3.4}$$

and ϕ is some constant phase.

Now, we can apply the standard method for deriving the separatrix map and the layer width (see Refs. 1 and 13 for details):

$$w_s = \Delta H_0 / K \approx 4\pi f \lambda_0^2 \exp(-\pi\lambda_0/2), \tag{3.5}$$

where ΔH_0 is the layer width in energy, $\lambda_0 = \omega_d/\omega_g = \lambda\omega_d/\sqrt{2}$, and $\lambda \equiv 1/\sqrt{K}$. Besides the usual approximations for such evaluations, an additional factor $f \sim 1$ shows up for the model (2.1) because the relative perturbation $|V/H_0| \sim 1$ is not small. In the particular case $N=1$, which reduces to the well studied standard map, this factor $f \approx 2.15$ was found in numerical experiments,[1] and later confirmed with much better accuracy in Ref. 20: $f = 2.255...$. The best theoretical value recently derived is $f \approx 2.14$ (Ref. 21). Uncertainty in this factor limits the theoretical accuracy of relation (3.5). It is partly balanced by an underestimated layer width, and also by a factor of 2 as discussed above.[14] Hence the factor $f = f_{th}/f_n$ in Eq. (3.5) is actually the ratio of a theoretical f_{th} to the correction $f_n = w_\infty / w_s$ of the empirical w_s' value (for 100 oscillation periods in our case) to obtain the true value w_∞ for infinitely many periods.

In this small-λ region the width w_s does not depend on N (Fig. 1) because the original interaction is local. However, this region is rather narrow. A comparison of numerical data for $L=2$ with theory (3.5) is presented in Fig. 1 (the dotted line a). The value of $f=0.64$ was obtained from the three leftmost points in Fig. 1 ($\ln\lambda=1.5-2.5$) with rms deviation from the theory (3.5) $\Delta \ln w_s=\pm 0.53$. Assuming the empirical correction[14] $f_n=2$ gives $f_{\rm th}=1.3$, which is rather different from that in the standard map.

4. LARGE-λ LIMIT: STATISTICAL ESTIMATES

For large λ the layer width, as well as the AD rate, progressively exceeds the simple estimate (3.5) (Fig. 1). This was noticed already in the first numerical experiments on AD.[1] Evidently this effect, which is somewhat strange at first glance, is due to higher-harmonic driving resonances, even though they are much weaker. Generally, such resonances are present in the original Hamiltonian (1.1), and their amplitudes V_{nm} are explicitly given. However, in the model (2.1) under consideration here this is not the case, and the higher perturbation harmonics show up only in higher orders of the perturbation expansion with respect to small perturbation parameter $K\ll 1$. The mechanism for generating higher-harmonic terms is related to the modulation of each unperturbed frequency p_i by any other degree of freedom. In particular, this general mechanism transforms the original local interaction between degrees of freedom in the system into a global one. Approximately, the higher-order amplitudes $V_n\sim K^n=\exp(n\ln K)$, and their decay rate σ (per freedom) can be assumed in the form[17]

$$\sigma=\ln(A/K) \tag{4.1}$$

with some constant A depending on a particular shape of the perturbation. In our model (2.1) the leading higher terms roughly correspond to $A\sim 2$, which we will use below. Notice that the amplitudes do not depend on the external perturbation harmonic m, since it is a δ-function.

A counterbalance to the weaker higher perturbation terms is the smaller λ (1.5) due to the smaller detuning ω_{nm} (1.2). Generally, the dependence $\omega_{nm}(n,m,\omega)$ is very complicated, with wild fluctuations, and exact evaluation of a higher-order perturbation is practically impossible and even useless beyond a few first terms.[21] However, the leading dependence can be found as follows (see, e.g., Refs. 22 and 23 and also Refs. 1, 15, and 17):

$$\omega_{nm}=\frac{\Omega}{q^{L-1}}\,F_{nm}(\omega), \tag{4.2}$$

where $q=\langle|n_i|\rangle$ is average absolute value of the components of integer vector n and now the new function F_{nm} describes the fluctuations only. The latter are quite big, which is the main obstacle for reliable estimates. In some special cases the function $F_{nm}=F_0$ is simply a constant. For example, for the case $L=2$ and frequency ratio $R=\omega/\Omega=(\sqrt{5}-1)/2$ (''the most irrational'' real number) we have $1/F_0=R+1/R=\sqrt{5}$. Generally, only a sort of statistical estimate can be obtained by setting $F_{nm}(\omega)\approx F_f\approx$const to some average value to be fitted from numerical data.

Now a particular term of the higher-order perturbation takes a form similar to Eq. (3.4):

$$V_n\sim\exp(-q\sigma(L-1))\cos\left(\frac{q\psi_1}{2}-\omega_{nm}t+\phi_{nm}\right), \tag{4.3}$$

where the factor $L-1$ is less by 1 than the full number of frequencies because of the δ-function in the Hamiltonian (2.1), as discussed above. Assuming again that the term (4.3) provides the main contribution to the formation of the chaotic layer, which seems to be plausible owing to the big detuning fluctuations, we arrive, analogously to Eq. (3.5), at the following estimate for the layer width:

$$w_s\sim(2e\lambda_n/q)^q\,\exp(-E(n)). \tag{4.4}$$

Here the principal exponent is (cf. Eq. (1.7))

$$E(n)=q\sigma(L-1)+\frac{\pi\lambda_n}{2},\quad \lambda_n=\frac{\omega_{nm}}{\omega_g}\approx\lambda_0\,\frac{F_f}{q^{L-1}}, \tag{4.5}$$

where $\lambda_0=\Omega/\omega_g=\lambda\Omega/\sqrt{2}$, and $\lambda\equiv 1/\sqrt{K}$ (Fig. 1).

The minimum of $E(n)$ is ($\Omega=2\pi$)

$$E_{\min}=\sigma^p L\Lambda^{1/L},\quad \Lambda=\frac{\pi^2}{\sqrt{2}}\,F_f\lambda,\quad p=1-\frac{1}{L}, \tag{4.6}$$

and is reached at $q\approx q_0$, where

$$q_0^L\approx\frac{\Lambda}{\sigma},\quad \frac{\lambda_n}{q_0}\approx\frac{2\sigma}{\pi}. \tag{4.7}$$

The latter relation shows that the factor (λ_n/q) in Eq. (4.4) approximately reduces to a constant $\sigma\to\sigma_L$ which renormalizes the amplitude decay rate, where

$$(L-1)\sigma_L\approx(L-1)\sigma-\ln\,\sigma-\ln\frac{4}{\pi}-1>0. \tag{4.8}$$

The latter inequality is a necessary condition for the validity of these approximate relations. This condition is satisfied for sufficiently large original σ, or small K (see Eq. (4.1)).

Finally, the approximate relation for the layer width in this limit reads

$$\ln w_s\approx A_f-b(L)\sigma_L^p L\Lambda^{1/L}. \tag{4.9}$$

This theoretical dependence is also shown in Fig. 1 (curve b_2) for $L=2$ and fitted values $A_f=5.42$, and $F_f=0.34$ for the detuning parameter in Eq. (4.6). The rms deviation for 5 points ($\ln\lambda=2-4$) is $\Delta\ln w_s=\pm 0.71$. While the average detuning F_f has a reasonable value, the factor A_f seems too big (see next section). Apparently, this discrepancy characterizes the accuracy of our statistical estimates. The additional parameter $b(L)=1$ was set equal to unity for $L=2$, and will be discussed in detail in Sec. 5 below.

For bigger L the behavior is completely different, and this, our most interesting result, will be described in the next section.

5. INTERMEDIATE ASYMPTOTICS: FAST ARNOLD DIFFUSION

The crucial change in the dependence $w_s(\lambda)$ stems from the factor $L-1$ in the expression for the exponent $E(n)$ (4.5). The effect of this factor was previously missed in Refs.

1 and 5 (cf. Eq. (1.7)). Indeed, it leads to a nonmonotonic dependence $w_s(L)$ according to Eq. (4.9). The latter was derived from optimization with respect to the average harmonic number q among the driving resonances with the maximum dimension $L=N$ only (see Eq. (4.2)). Meanwhile, there are also resonances of lower dimension with $\tilde{L}<L$. Hence we need a second optimization, now with respect to $\tilde{L}$, as was first done in Ref. 14 (see also Ref. 17). First, we explain the idea of optimization for a simple example (cf. Eq. (4.9))

$$w_s=\exp(-L\lambda^{1/L}). \tag{5.1}$$

The new factor L decreases the layer width as L grows, and thus counteracts the increase in w_s due to the dependence $\lambda^{1/L}$. For any pair $L_1<L_2$ there is a certain value of $\lambda=\lambda^*$ at which both w_s values coincide,

$$\lambda^*=(L_2/L_1)^{L_1L_2/(L_2-L_1)}. \tag{5.2}$$

For $\lambda<\lambda^*$ we have $w_s(L_1)>w_s(L_2)$ and vice versa. Thus, for a given λ the particular $\tilde{L}(\lambda)$ should be found which maximizes w_s. In this way we would obtain a broken line which is the envelope of the family of curves $w_s(\lambda,\tilde{L})$. Interestingly, the existence of such a family of intersecting curves could already be inferred (but was missed) from the validity of approximation (3.5) which corresponds to $\tilde{L}=1$ (Refs. 1, 2, 6, and 7).

For $L\gg 1$ a smooth approximation to the envelope is found from the local condition

$$\frac{dw_s}{d\tilde{L}}=-w_s\lambda^{1/\tilde{L}}\left(1-\frac{\ln\lambda}{\tilde{L}}\right)=0, \tag{5.3}$$

whence we obtain the optimal value

$$\tilde{L}_0(\lambda)=\ln\lambda \tag{5.4}$$

and

$$w_{\max}(\lambda)=w_s(\tilde{L}_0)=\lambda^{-e}, \tag{5.5}$$

where $e=\exp(1)$. Thus, the dependence of the layer width on the adiabaticity parameter becomes a power law, provided that $\tilde{L}_0\leqslant L$, or

$$\lambda\leqslant\lambda_L=e^L, \tag{5.6}$$

i.e., for a not-too-weak perturbation. This border is, of course, much higher (in ε) than that in the rigorous theory (cf. Eq. (1.11)). We term (5.6) the intermediate asymptotic region, as contrasted to the far asymptotic limit for the reversed inequality. The former is always bounded from above but rapidly grows with L, and may be arbitrarily large as $L\to\infty$.

We call this regime fast Arnold diffusion (FAD). Within the domain (5.6) the layer width (and diffusion rate) does not depend on L, but for any fixed L and $\lambda\to\infty$ the Nekhoroshev-like dependence (4.9) is recovered asymptotically.

In Fig. 3 the power-law mechanism is illustrated, for the simple example (5.1), by plotting the family of curves $\ln(w_s(\lambda,\tilde{L})/w_{\max})$ which are tangent to the line of maximal $w_{\max}(\lambda)$ (5.5) up to the largest $\tilde{L}=L=5$.

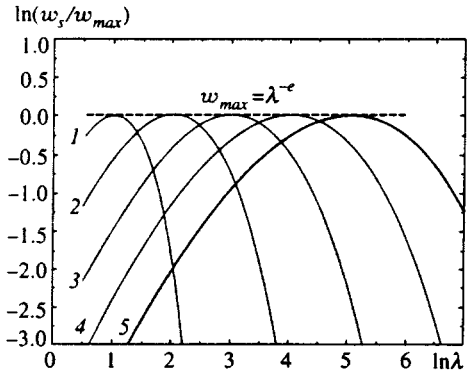

FIG. 3. A scheme of the family $w_s(\lambda,\tilde{L})$, for $\tilde{L}=1-5$ as indicated, with maximal $\tilde{L}=L=5$ which form the smooth power-law dependence (5.5) shown by dotted straight line.

For the more realistic asymptotic relation (4.9) the optimization is more complicated because of the additional dependence on L via σ_L^p. That can be partly removed by approximate renormalization: $\Lambda_0\to\Lambda_0/\sigma$. For $L\gg 1$ the remaining dependence (4.8) is weak and can be neglected, at least in evaluating the optimal $\tilde{L}_0$, which now becomes (cf. Eq. (5.4))

$$\tilde{L}_0(\lambda)\approx\ln(\Lambda/\sigma). \tag{5.7}$$

However, we retain the more accurate value of σ_L (4.8) in the final expression:

$$\ln w_s\approx A_f-b_fe\left[\sigma\ln\left(\frac{\Lambda}{\sigma}\right)-\ln\left(\frac{4\sigma}{\pi}\right)-1\right], \tag{5.8}$$

which is the main result of our studies. It is compared with numerical data in Fig. 1 (curve c, see also Fig. 2). Besides two fitting parameters previously used in Eq. (4.9) (curve b_2 in Fig. 1), which now take somewhat different values; ($A_f=-1.05$ and $F_f=0.4$), we have to introduce a third one, $b_f=0.29$. The fitting of empirical data has been performed for $N=5$, 7, 9, 15 only. We excluded data for $N=3$, 4 as they seem to violate the condition (cf. Eq. (5.6))

$$\Lambda\leqslant\Lambda_L=\frac{\pi^2}{\sqrt{2}}F_f\lambda_L\approx\sigma e^L \tag{5.9}$$

for $\ln\lambda\gtrsim 5$ (see Figs. 1 and 2). Using the above fitted value for $F_f=0.4$, and Eq. (4.1) for $\sigma=\ln(2/K)=\ln(2\lambda^2)$ we obtain from Eq. (5.9) $\ln\lambda_3\approx 4.2$ and $\ln\lambda_4\approx 5.5$. While the first value is close to the empirical one, the second is too large. The origin of this discrepancy is not completely clear, but it may be caused by fluctuations. Apparently, the latter are mainly related to the detuning function $F_{nm}(\omega)$ which fluctuates with both the harmonic numbers and the set of frequencies for different L. Interestingly, while the optimal harmonic number q_0 increases with $\lambda>\lambda_L$ as in (4.7), it remains approximately constant,

$$q_0\approx e\approx 3, \tag{5.10}$$

in the whole FAD region (5.9). This follows directly from Eqs. (4.7) and (5.7). Surprisingly, the above asymptotic relations remain reasonably good in spite of the relatively

small q_0 value (Figs. 1 and 2). Notice, however, that the number of resonances $\sim q_0^{\widetilde{L}_0}=\Lambda/\sigma$ still increases with λ.

Detuning fluctuations in F_f were calculated from the numerical data using the relation (see Eq. (5.8))

$$\frac{d\ \ln w_s}{d\ \ln F_f}=-b_f e\sigma_L\approx -b_f e(0.7+2\ \ln\lambda), \tag{5.11}$$

which gives for the rms dispersion

$$\langle\Delta\ \ln F_f\rangle_4=0.18, \quad \text{and } \langle\Delta\ \ln F_f\rangle_6=0.25. \tag{5.12}$$

The first value is the average over 4 cases with $N=5, 7, 9, 15$ as in the main fitting; for the second $N=3, 4$ are also included. The latter value is used in Fig. 2 for rms fluctuations $\Delta\ \ln w_s$ according to Eq. (5.11).

The accuracy of our theory does not allow for a reasonable estimate of the factor $A_f\approx -1$ in the main relation (5.8), whose value is considerably smaller than $A_f\approx 5$ in Eq. (4.9). However, the value of the new fitting parameter $b_f=0.29$, which we had to introduce in Eq. (5.8) instead of $b(2)=1$ in Eq. (4.9), is a problem for the theory. It is impossible to fit the data for large L with the latter value or vice versa, i.e., with $b(2)=0.3$, as in Eq. (5.8) except for $L=2$, unless one assumes the value $F_f=3$ in Eq. (4.9) instead of 0.3, which seems too big. In any event, something happens in going from $L=2$ to $L\geqslant 3$, which is obvious from the data in Fig. 1. To reconcile these data with the above theory one needs to assume a drop either in the parameter b from 1 to 0.3 (with approximately the same $F_f\approx 0.4$) or in the parameter F_f from 3 to 0.4 (with approximately the same $b\approx 0.3$ still to be explained anyway). Actually, the value $F_f=3$ for $L=2$ would contradict the rigorous upper bound $F_f\leqslant 1$.[22] So we have to understand the first possibility above.

In Ref. 17, using a somewhat different approach, the following expression has been derived for the parameter b in the relation (2.11), similar to Eq. (5.8) above: $b\approx 1/\pi\sqrt{e}=0.19$. This value is close to the present empirical one, $b_f=0.29$. However, the former does not fit the far asymptotic expression (4.9) for $L=2$, as discussed above.

A qualitative explanation of the decrease in $b(L)$ with L could be related to an underestimate of the perturbation Fourier amplitudes in Eq. (4.3). Indeed, we assumed that the amplitudes decay independently for each degree of freedom (factor $L-1$). However, the higher harmonics may arise in the perturbation series not individually but in groups, thus decreasing the effective parameter L or σ. The former possibility is excluded by the assumed expression (4.2) for detuning. Hence we guess the effective amplitude decay rate in the form $\sigma\to b\sigma$ with empirical $b\approx b_f\approx 0.3$.

A different value of $b=1$ for $L=2$ is also explained in this way because in that case only a single oscillation frequency remains. However, another important question arises: is the new factor $b(L)$ a constant for $L\geqslant 3$ or does it change still further with L? In other words, is FAD really independent of N? Our empirical data seem to confirm such independence. Even though there are quite big fluctuations for large λ they do not reveal any systematic variation of w_s with L. This is especially clear from Fig. 2 where the difference between the numerical data and the theory is shown. Moreover, the theory explains even a small dip in the dependence $w_s(\lambda)$ around $\ln\lambda=3$. This results from a deviation of the approximate smoothed envelope (5.8) from the family of curves $w_s(\lambda,\widetilde{L})$, three of which are shown in Fig. 2 (for $\widetilde{L}=2, 3, 4$, cf. Fig. 3) as calculated from Eq. (4.9) with the factor $b(2)=1$, and $b(3)=0.29$.

If the above hypothesis is true a new fitting is required, because the renormalization $\sigma\to b\sigma$ would result in more complicated expressions than just a single factor in Eqs. (5.8) and (4.9). By doing so we have found that $w_s(\lambda)$ according to Eq. (5.8) changed negligibly after some changes in the fitting parameters: $A_f=-0.88$, $b_f=0.28$, $F_f=0.21$ which appear to be reasonable also. A larger change $\delta A_f\approx 1$ occurs in the family of curves Eq. (4.9) for $L>2$, and their agreement with the smooth envelope (5.8) worsens owing to the approximate relation (5.7). To keep the above estimates more self-consistent we neglect all these minor changes, and retain the above relations with a single parameter $b_f=0.29$ for $L>2$. In any event, the relations, which are approximate anyway, are much simpler in this form.

Interestingly, half of the data in Fig. 1 ($\ln\lambda\leqslant 4$, $L>2$) also fit a simple power law with exponent 6.3, which is very close to the value 6.6 obtained in Refs. 16 and 17 around $\ln\lambda\approx 2$. However, for larger $\ln\lambda>4$ the deviation from such a simple dependence (it would be a straight line in Fig. 1) progressively increases in accordance with the theory (5.8).

6. DISCUSSION

We have performed detailed investigations into fast Arnold diffusion, a new regime of AD when the diffusion rate depends on the perturbation strength $\varepsilon=K$, for the models (1.1) and (2.1) respectively, or on the adiabaticity parameter $\lambda\sim 1/\sqrt{\varepsilon}\sim 1/\sqrt{K}$ as a power law (5.8) rather than an exponential like Eq. (4.9).

We made use of a specific model (2.1) which is relatively simple and very convenient for numerical experiments with arbitrary number of degrees of freedom N but, at the same time, is rather difficult for theoretical analysis. This is because the model represents the limiting case of the local interaction between degrees of freedom. Not only between two degrees of freedom, which would model a pair interaction in a broad class of physical systems, but even further restricted to the coupling between two nearest-neighbor degrees of freedom in a chain. Moreover, the coupling is harmonic, so that only three-frequency primary resonances (for the two degrees of freedom and for the driving perturbation) with harmonic numbers $n=\pm 1$ show up in the original Hamiltonian (2.1) independent of N. As a result, the higher-harmonic multifrequency resonances, which make the principal contribution to AD, arise only in higher-order perturbation terms, which makes the theory very difficult from the beginning. We circumvented this difficulty by a plausible and simple conjecture (4.1) for the decay rate of the high-order perturbation amplitudes. However, to reconcile the empirical data with the theory we had, later on, to further modify this conjecture by introducing the additional factor $b(L)$ into our main relations (4.9) and (5.8). Even though we suggest in section 5 a qualitative explanation for $b(L)\neq 1$,

the origin of this additional dependence is not yet completely clear, and it constitutes an open question in our theory. Apparently, this is related to the specific Hamiltonian (2.1) as discussed above.

The factor $b=0.29$, assumed to be constant for $L=N>2$, is one of the three fitting parameters in our main theoretical relation (5.8) for the FAD. As explained in section 2, we actually computed and calculated the chaotic layer width w_s related to the diffusion rate via estimate (1.6). The second fitting parameter $F_f=0.4$, which describes the detuning fluctuations w_{nm} (1.2), also cannot be evaluated in the present state of the theory but was found numerically to have a plausible value. Finally, the third fitting parameter A_f remains completely out of theoretical reach and simply characterizes the global accuracy of the theory. We recall that all our estimates except the simplest one (3.5) are of a statistical nature, owing to the large detuning fluctuations. Within this accuracy and fluctuations, the agreement between the empirical data and the theory as presented in Figs. 1 and 2 can be regarded as satisfactory, especially taking into account the big range of w_s variation, almost 22 orders of magnitude!

Surprisingly, all this huge range corresponds to the intermediate asymptotic region ($1\ll\lambda\ll\lambda_L$, see Eq. (5.9)) with FAD, starting even at a relatively small $L=N\geqslant 5$. Even for $L=3$ and 4 the FAD range is apparently of a comparable size, and only for the minimal $L=2$ does the far (exponential) asymptotic ($\lambda\gg\lambda_L$) behavior clearly show up. As already discussed in section 5, the sharp change in $w_s(\lambda)$ from $L=2$ to $L=3$ is precisely due to the ''mysterious'' parameter b, which drops by a factor of 3. Unfortunately, this does not allow us to reach the far asymptotic limit and to confirm the exponential dependence (4.9) on N for $\lambda>\lambda_L$ beyond the minimum $L=2$. Meanwhile, this would be important to decide on the different interpretation of N-independent diffusion for large N in Ref. 16. The authors of the latter conjectured that this independence is a result of the local interaction in the model (2.1). This contradicts our theory but not as yet the direct empirical evidence. At the moment we can only remark that their reference to Wayne's theory[24] for the same model is irrelevant. Indeed, Wayne proved a long N-independent stability for very special, nonresonant, initial conditions (theorem 1.1), whereas AD occurs within chaotic layers only, i.e., also for highly specific but resonant initial conditions. Thus, the former theory is related to a global chaos rather than to KAM integrability with its peculiar Arnold web of chaotic layers.

In the case of a global interaction (1.1) with strong nonlinearity (1.4) and uniform amplitude decay rate our theory remains valid, and even becomes simpler as $\sigma=\text{const}$. However, the numerical experiments would be much more difficult for large N. On the other hand, both the FAD range (5.9) and the diffusion rate there depend generally on the number of incommensurate unperturbed frequencies $L=N+M-r$ (1.9), which may be large if M, the number of driving perturbation frequencies, is large.

Fast Arnold diffusion should not be confused with the much faster diffusion in degenerate systems or those with nonconvex energy surfaces (Sec. 1). In the latter case the diffusion mechanism is completely different. Apparently, this sort of diffusion was recently observed in numerical experiments with the classical model of the hydrogen atom in crossed electric and magnetic fields.[25]

In the present study we have chosen one of the strongest primary resonances as guiding, with amplitude $V_g=V_1\sim K$ (Sec. 2). In case of a high-harmonic guiding resonance ($V_g=V_n$, $n\gg 1$) the main effect would be a tremendous drop in the diffusion rate due to the exponential rise of the adiabaticity parameter with q (see Eq. (4.5)):

$$\lambda_n\sim\exp\left(\frac{\sigma}{2}Lq\right)\sim\exp\left(\frac{\sigma}{2}L\rho^{1/L}\right), \tag{6.1}$$

where $\rho(n)\sim q^L$ is the density of the guiding resonances in the Arnold web with harmonic numbers up to q (cf. Eq. (4.2)). Hence, the diffusion rate in the intermediate asymptotic region drops exponentially with q or ρ, Eq. (5.8), and even as a double exponential in the far asymptotic limit!

In conclusion, our present studies confirm the previous conjecture and preliminary empirical data[14,17] concerning a new regime of fast Arnold diffusion. Moreover, we have found that in multifrequency systems ($L\gg 1$), in particular, large ones ($N\gg 1$), the FAD range in the perturbation (5.9) is fairly big, so that this regime appears to be typical, in a sense, and might be important in various applications.

We are very grateful to our colleagues in the University of Milano at Como for the hospitality during the essential part of this study. This work was partially supported also by the Russia Fund for Fundamental Research (Grant 95-01-01047).

*E-mail: chirikov@inp.nsk.su

[1] B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
[2] A. Lichtenberg and M. Lieberman, *Regular and Chaotic Dynamics*, Springer (1992).
[3] V. I. Arnold, Dokl. Akad. Nauk SSSR **156**, 9 (1964).
[4] B. V. Chirikov, *Studies in the Theory of Nonlinear Resonance and Stochasticity*, preprint INP-267, Novosibirsk (1969) (Engl. trans., CERN Trans. 71–40, 1971).
[5] G. V. Gadiyak, F. M. Izrailev, and B. V. Chirikov, *Proc. 7th Intern. Conf. on Nonlinear Oscillations*, Akademie-Verlag, Berlin (1977), Vol. **II, 1. p.** 315.
[6] J. Tennyson, M. Lieberman, and A. Lichtenberg, AIP Conf. Proc. **57**, 272 (1979); M. Lieberman, Ann. N.Y. Acad. Sci. **357**, 119 (1980).
[7] B. V. Chirikov, J. Ford, and F. Vivaldi, AIP Conf. Proc. **57**, 323 (1979).
[8] T. Petrosky, Phys. Rev. A **29**, 2078 (1984).
[9] B. Wood, A. Lichtenberg, and M. Lieberman, Physica D **71**, 132 (1994).
[10] N. N. Nekhoroshev, Usp. Mat. Nauk **32**, (6), 5 (1977).
[11] P. Lochak, Phys. Lett. A **143**, 39 (1990); Usp. Mat. Nauk **47**, (6), 59 (1992); P. Lochak and A. Neishtadt, Chaos **2**, 495 (1992).
[12] A. A. Chernikov, R. Z. Sagdeev, and G. M. Zaslavsky, Physica D **33**, 65 (1988); A. Lichtenberg and B. Wood, Phys. Rev. A **39**, 2153 (1989); B. V. Chirikov and V. V. Vecheslavov, in *From Phase Transitions to Chaos*, ed. by G. Györgyi *et al.* World Scientific (1992), p. 273.
[13] N. N. Filonenko, R. Z. Sagdeev, and G. M. Zaslavsky, Nucl. Fusion **7**, 253 (1967); G. M. Zaslavsky and N. N. Filonenko, Zh. Eksp. Teor. Fiz. **54**, 1590 (1968); G. M. Zaslavsky and B. V. Chirikov, Usp. Fiz. Nauk **105**, 3 (1971); A. Rechester and T. Stix, Phys. Rev. Lett. **36**, 587 (1976); D. Escande, Phys. Rep. **121**, 163 (1985); V. F. Lazutkin, Dokl. Akad. Nauk SSSR **313**, 268 (1990); V. V. Afanasiev *et al.*, Phys. Lett. A **144**, 229 (1990).
[14] B. V. Chirikov and V. V. Vecheslavov, *How Fast is Arnold Diffusion?* Preprint INP 89–72, Novosibirsk (1989). B. V. Chirikov and V. V. Vecheslavov, in *Analysis etc.* ed. by P. Rabinowitz and E. Zehnder, Academic Press (1990), p. 219.

[15] B. V. Chirikov, Fiz. Plasmy **4**, 521 (1978); Proc. R. Soc. London, Ser. A **413**, 145 (1987).
[16] K. Kaneko and T. Konishi, Phys. Rev. A **40**, 6130 (1989); T. Konishi, Prog. Theor. Phys. Suppl. **98**, 19 (1989); T. Konishi and K. Kaneko, J. Phys. A **23**, L715 (1990).
[17] B. V. Chirikov and V. V. Vecheslavov, J. Stat. Phys. **71**, 243 (1993).
[18] *Resonances in the Motion of Planets, Satellites and Asteroids*, ed. by S. Ferraz-Mello and W. Sessin, Univ. de Sao Paulo, Brazil (1985); J. Wisdom, Icarus **72**, 241 (1987); J. Laskar, Nature (London) **338**, 237 (1989); Icarus **88**, 266 (1990); Astron. Astrophys. **287**, L9 (1994).
[19] B. V. Chirikov, Priroda **7**, 15 (1982); *Nolinear Dynamics Aspects of Particle Accelerators*, Lecture Notes in Physics, Vol. 247, Springer (1986);A. Lichtenberg, Phys. Fluids B **4**, 3132 (1992).
[20] V. F. Lazutkin *et al.*, Physica D **40**, 235 (1989).
[21] V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **109**, 2208 (1996) [JETP **82**, 1190 (1996)].
[22] A. Ya. Khinchin, *Continued fractions*, Fizmatgiz, Moskva (1961) (in Russian); J. Cassels, *An Introduction to Diophantine Approximation*, Cambridge University Press (1957); W. Schmidt, *Diophantine Approximations*, Lecture Notes in Mathematics **785**, Springer (1980).
[23] B. V. Chirikov, Chaos, Solitons and Fractals **1**, 79 (1991).
[24] E. Wayne, Commun. Math. Phys. **104**, 21 (1986).
[25] J. von Milczewski, G. Diercksen, and T. Uzer, Phys. Rev. Lett. **76**, 2890 (1996).

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

Erratum: Acoustics of a relativistic superfluid [JETP 84, 729–733 (1997)]

Yu. V. Vlasov

L. D. Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117334 Moscow, Russia
Zh. Éksp. Teor. Fiz. **112**, 1147 (September 1997)

[S1063-7761(97)02509-2]

Through an oversight on the part of the author, there was no indication that the article, as a whole, describes the propagation of shocks and sound waves in a relativistic two-component medium in which the components interact. Not all the results apply specifically to a relativistic superfluid, but only the equations in the acoustic limit and the derivation of the formulas for the velocity of first and second sound (Secs. 4–6).

 1063-7761/97/090625-01$10.00